सत्यमेव जयते

INDIAN AGRICULTURAL
RESEARCH INSTITUTE, NEW DELHI.

I. A. R. I. 6.
MGIPC—S1—6 AR/54—7-7-54—10,000.

TRANSACTIONS OF

THE ROYAL SOCIETY OF CANADA

SECTION III

CHEMICAL, MATHEMATICAL, AND PHYSICAL SCIENCES

THIRD SERIES VOLUME XXXVII—SECTION III

MAY, 1943

56667

Lending Library
Imperial Agricultural Research Institute,
New Delhi

OTTAWA
THE ROYAL SOCIETY OF CANADA
1943

TABLE OF CONTENTS

Periodic Orbits for Four Finite Bodies with Repulsive and Attractive Forces. By DANIEL BUCHANAN, F.R.S.C. 1

An Arithmetical Identity. By WALTER H. GAGE 9

A Short Table of the Toronto Function. By A. H. HEATLEY ... 13

Entire Functions Bounded on a Set. By M. S. MACPHAIL 31

The Colours of Snowfields in Sunlight. By W. E. KNOWLES MIDDLETON, F.R.S.C 39

On a Statistical Model of Isotropic Turbulence. By J. L. SYNGE, F.R.S.C., and C. C. LIN 45

Transactions of the Royal Society of Canada
SECTION III

SERIES III MAY, 1943 VOL. XXXVII

PERIODIC ORBITS FOR FOUR FINITE BODIES WITH REPULSIVE AND ATTRACTIVE FORCES

By DANIEL BUCHANAN, F.R.S.C.

1. INTRODUCTION

THIS paper deals with periodic restricted motion of four finite bodies of equal masses having attractive and repulsive forces. The system consists of two pairs of *like* bodies or, alternately, two pairs of *unlike* bodies. The like bodies repel and the unlike attract. The forces of repulsion and attraction vary according to the Newtonian law of the inverse square and their magnitudes are in a constant ratio.

One pair of *like* bodies is constrained to move in a plane while the bodies of the other *like* pair move to and fro on opposite sides of this plane and along its normal.

There are three classes of orbits differing according to their periods. The periods vary with the ratio of the magnitudes of the forces of attraction and repulsion.

2. THE DIFFERENTIAL EQUATIONS OF MOTION

Let the planar motion be in the xy-plane and the linear motion be in the z-axis. The coordinates of the bodies will be designated as follows:

Pair *one-two*; I $(x_1, y_1, 0)$, II $(x_2, y_2, 0)$,
Pair *three-four*; III $(0, 0, z_3)$, IV $(0, 0, z_4)$.

The motion is restricted so that

$$x_2 = -x_1,\ y_2 = -y_1,\ z_4 = -z_3. \tag{1}$$

Let the units of time and space be chosen so that the gravitational constant of attraction is unity and let k^2 denote the ratio of the magnitude of repulsion to that of attraction. Let the mass of each body be unity. Then the differential equations of motion for I and III are

$$\begin{aligned} x_1'' &= -\frac{2x_1}{r_{13}^3} + \frac{k^2(x_1 - x_2)}{r_{12}^3}, \\ y_1'' &= -\frac{2y_1}{r_{13}^3} + \frac{k^2(y_1 - y_2)}{r_{12}^3}, \\ z_3'' &= -\frac{2z_3}{r_{13}^3} + \frac{k^2(z_3 - z_4)}{r_{34}^3}, \\ r_{13}^2 &= x_1^2 + y_1^2 + z_3^2, \\ r_{12}^2 &= (x_1 - x_2)^2 + (y_1 - y_2)^2, \\ r_{34}^2 &= (z_3 - z_4)^2. \end{aligned} \tag{2}$$

When restrictions (1) are applied, the subscripts 1, 2, 3, dropped and the motion is transformed to cylindrical coordinates by putting

$$x = r\cos\theta,\ y = r\sin\theta,$$

the equations of motion become

$$\left.\begin{aligned} r'' - r\theta'^2 &= -\frac{2r}{(r^2+z^2)^{3/2}} + \frac{k^2}{4r^2}, \\ r\theta'' + 2r'\theta' &= 0, \text{ or } r^2\theta' = c, \text{ a constant}, \\ z'' &= -\frac{2z}{(r^2+z^2)^{3/2}} + \frac{k^2}{4z^2}. \end{aligned}\right\} \tag{3}$$

When θ' is eliminated we obtain

$$\left.\begin{aligned} r'' &= \frac{c^2}{r^3} - \frac{2r}{(r^2+z^2)^{3/2}} + \frac{k^2}{4r^2}, \\ z'' &= -\frac{2z}{(r^2+z^2)^{3/2}} + \frac{k^2}{4z^2}. \end{aligned}\right\} \tag{4}$$

A particular solution of (4) is

$$\left.\begin{aligned} r &= r_0 = \frac{4c^2}{(4 - k^{4/3})^{3/2} - k^2}, \\ z &= z_0 = \mu r_0, \\ \mu &= k^{2/3}/(4 - k^{4/3})^{1/2}. \end{aligned}\right\} \tag{5}$$

Since r_0 must be positive, then $k^2 < 2\sqrt{2}$.

3. THE EQUATIONS OF DISPLACEMENT

Let

$$\left.\begin{aligned} r &= r_0(1 + \epsilon\rho), \\ z &= \mu r_0(1 + \epsilon\zeta), \\ t - t_0 &= \sqrt{1 + \delta}\,\tau, \end{aligned}\right\} \tag{6}$$

where ϵ and δ are parameters, δ vanishing with ϵ, and τ is the new independent variable. Then equations (4) become, after dividing through by ϵr_0 or $\epsilon\mu r_0$ $(\neq 0)$

$$\left.\begin{aligned} &\{D^2 + \alpha_1(1+\delta)\}\rho + \alpha_2(1+\delta)\zeta = \\ &\qquad (1+\delta)(\epsilon R_2 + \epsilon^2 R_3 + \ldots + \epsilon^j R_{j+1} + \ldots), \\ &\alpha_3(1+\delta)\rho + \{D^2 + \alpha_1(1+\delta)\}\zeta = \\ &\qquad (1+\delta)(\epsilon Z_2 + \epsilon^2 Z_3 + \ldots + \epsilon^j Z_{j+1} + \ldots), \\ &D = \frac{d}{d\tau}, \end{aligned}\right\} \quad (7)$$

where the right sides are power series in ϵ converging for ϵ sufficiently small numerically. The R's and Z's are homogeneous polynomials in ρ and ζ of the degree of the subscript. The values of the α's are

$$\begin{aligned} \alpha_1 &= \frac{2}{r_0^3}\left[\frac{1+4\mu^2}{(1+\mu^2)^{5/2}} - \frac{\mu^3}{(1+\mu^2)^{3/2}}\right], \\ \alpha_3 - \alpha_1 &= -\frac{6}{r_0^3(1+\mu^2)^{5/2}}, \\ \alpha_2 &= \mu^2\alpha_3. \end{aligned} \quad (8)$$

1. The Equations of Variation

If we consider only the terms of (7) that are independent of ϵ we obtain the equations of variation

$$\begin{aligned} (D^2 + \alpha_1)\,\rho + \alpha_2\zeta &= 0, \\ \alpha_3\,\rho + (D^2 - \alpha_3)\zeta &= 0. \end{aligned} \quad (9)$$

The indicial equation is

$$D^4 + (\alpha_1 - \alpha_3)D^2 - \alpha_3(\alpha_1 + \alpha_2) = 0.$$

The roots are

$$D^2 = \frac{(\alpha_3 - \alpha_1) \pm \sqrt{(\alpha_1 - \alpha_3)^2 + 4\alpha_3(\alpha_1 + \alpha_2)}}{2} \quad (10)$$

or

$$= \frac{(\alpha_3 - \alpha_1) \pm \sqrt{(\alpha_1 + \alpha_3)^2 + 4\mu^2\alpha_3^2}}{2}. \quad (11)$$

The discriminant is positive and hence both values of D^2 are real.

We now show that both values of D^2 are negative. The expression $\alpha_1 - \alpha_3$ has the value

$$\frac{8 - 2\mu^3}{r_0^3(1+\mu^2)^{3/2}},$$

and μ lies in the range $0 < \mu < 1$. Hence $\alpha_1 - \alpha_3$ is positive.

Next we must know the sign of $4\alpha_3(\alpha_1 + \alpha_2)$ in (10). This expression is

$$-\frac{18(1-\mu^2)}{(1+\mu^2)^4}$$

and is always negative. Hence the radical in (10) yields a quantity which is numerically less than $\alpha_1 - \alpha_4$ and consequently both values for D^2 are negative. Suppose

$$D^2 = -\sigma_1^2 \text{ or } -\sigma_2^2.$$

Then the solutions of the equations of variation are

$$\begin{aligned}
\rho &= \lambda_1(A_1\cos\sigma_1\tau + A_2\sin\sigma_1\tau) + \lambda_2(A_3\cos\sigma_2\tau + A_4\sin\sigma_2\tau),\\
\zeta &= \quad A_1\cos\sigma_1\tau + A_2\sin\sigma_1\tau \;+\; \quad A_3\cos\sigma_2\tau + A_4\sin\sigma_2\tau,\\
\lambda_1 &= \frac{\alpha_2}{\alpha_1-\sigma_1^2} = -\frac{\alpha_3+\sigma_1^2}{\alpha_4}, \quad \lambda_2 = \frac{\alpha_2}{\alpha_1-\sigma_2^2} = -\frac{\alpha_3+\sigma_2^2}{\alpha_4}.
\end{aligned} \tag{12}$$

5. THE CONSTRUCTION OF THE SOLUTION

To integrate the differential equations (7) we substitute in (7)

$$\rho = \sum_{j=0}^{\infty} \rho_j\epsilon^j, \quad \zeta = \sum_{j=0}^{\infty} \zeta_j\epsilon^j, \quad \delta = \sum_{j=1}^{\infty} \delta_j\epsilon^j, \tag{13}$$

where

$$\rho_j = \rho_j(\tau), \quad \zeta_j = \zeta_j(\tau), \quad \delta_j = \text{constants}.$$

Let the resulting equations be cited by (7′). On equating the coefficients of the various powers of ϵ in (7′) we obtain sets of differential equations which define the various ρ_j and ζ_j. These equations may be integrated, as we shall show, and the various δ's and constants of integration at each step can be determined so that the ρ_j and ζ_j will be periodic and will satisfy certain initial conditions.

For initial values let us suppose $D\rho(0) = 0$, $\zeta(0) = 1$. Then

$$\begin{aligned}
&D\rho_j(0) = 0, \; j = 0, \ldots \infty,\\
&\zeta_0(0) = 1, \; \zeta_j(0) = 0, \; j = 1, \ldots \infty.
\end{aligned} \tag{14}$$

As ζ carries the factor ϵ in (6) we may choose $\zeta(0) = 1$ without loss of generality.

There are three types of solutions according to their periods.

Case I. Period $2\pi/\sigma_1$.

Terms in ϵ^0.

The terms in (7′) that are independent of ϵ are the same as (9). Hence the solutions which have the proper period are

$$\begin{aligned}\rho_0 &= \lambda_1(A_1^{(0)}\cos\sigma_1\tau + A_2^{(0)}\sin\sigma_1\tau),\\ \zeta_0 &= \quad A_1^{(0)}\cos\sigma_1\tau + A_2^{(0)}\sin\sigma_1\tau,\end{aligned} \tag{15}$$

where $A_1^{(0)}$, $A_2^{(0)}$ are arbitrary constants. When conditions (14) are imposed we obtain

$$\rho_0 = \lambda_1\cos\sigma_1\tau,\ \zeta_0 = \cos\sigma_1\tau.$$

Terms in ϵ

The differential equations at this step are

$$\left.\begin{aligned}(D^2+\alpha_1)\rho_1 + \alpha_2\zeta_1 = R^{(1)} &= -\delta_1(\alpha_1\rho_0+\alpha_2\zeta_0) + a_0^{(1)} + a_2^{(1)}\cos 2\sigma_1\tau,\\ \alpha_3\rho_1 + (D^2-\alpha_3)\zeta_1 = Z^{(1)} &= -\delta_1\alpha_3(\rho_0-\zeta_0) + b_0^{(1)} + b_2^{(1)}\cos 2\sigma_1\tau,\end{aligned}\right\}$$

where $a_0^{(1)}, \ldots, b_2^{(1)}$ are known constants.

In order that ρ_1 and ζ_1 may be periodic the coefficients of $\cos\sigma_1\tau$ in the equivalent expressions

$$(D^2-\alpha_3)R^{(1)} - \alpha_2 Z^{(1)} \text{ and}$$
$$-\alpha_3 R^{(1)} + (D^2+\alpha_1)Z^{(1)}$$

must vanish. From the second expression we obtain

$$\delta_1\alpha_3\,[\alpha_1+\alpha_2+\sigma_1^2] = 0.$$

Since $\alpha_1+\alpha_2+\sigma_1^2 \neq 0$ we must put $\delta_1 = 0$. Hence the desired solutions satisfying (14) are

$$\begin{aligned}\rho_1 &= A_0^{(1)} + A_1^{(1)}\cos\sigma_1\tau + A_2^{(1)}\cos 2\sigma_1\tau,\\ \zeta_1 &= B_0^{(1)} + B_1^{(1)}\cos\sigma_1\tau + B_2^{(1)}\cos 2\sigma_1\tau,\\ \delta_1 &= 0\end{aligned}$$

where the $A_0^{(1)}, \ldots, B_2^{(1)}$ are known constants.

Terms in ϵ^2.

From the coefficients of ϵ^2 in (7′) we obtain

$$\begin{aligned}(D^2+\alpha_1)\rho_2 + \alpha_2\zeta_2 &= R^{(2)},\\ \alpha_3\rho_2 + (D^2-\alpha_3)\zeta_2 &= Z^{(2)},\\ R^{(2)} &= -\delta_2(\alpha_1\rho_0+\alpha_2\zeta_0) + \sum_{j=0}^{3} a_j^{(2)}\cos j\sigma_1\tau,\\ Z^{(2)} &= -\delta_2\alpha_3(\rho_0-\zeta_0) + \sum_{j=0}^{3} b_j^{(2)}\cos j\sigma_1\tau,\end{aligned}$$

where $a_j^{(2)}$, $b_j^{(2)}$ are known constants. The undetermined δ_2 enters only as expressed above. As at the previous step the coefficient of $\cos\sigma_1\tau$ in

$$(D^2-\alpha_3)R^{(2)} - \alpha_2 Z^{(2)} \text{ and } -\alpha_3 R^{(2)} + (D^2+\alpha_1)Z^{(2)}$$

must vanish. We obtain

$$\delta_2\alpha_3(\alpha_1+\alpha_2+\sigma_1^2) = \text{a constant},$$

and therefore $\delta_2 = d_2$, a constant.

The particular integrals can be readily found. When the arbitrary constants of integration are determined so as to satisfy (14) we obtain

$$
\begin{aligned}
\rho_2 &= \sum_{j=0} A_j^{(2)} \cos j\sigma_1\tau, \\
\zeta_2 &= \sum_{j=0}^{3} B_j^{(2)} \cos j\sigma_1\tau, \\
\delta_2 &= d_2,
\end{aligned}
$$

where $A_j^{(2)}$, $B_j^{(2)}$, d_2 are known constants.

The remaining steps of the integration can be carried on in an entirely similar manner. The solutions at the general step are

$$
\begin{aligned}
\rho_n &= \sum_{j=0}^{n+1} A_j^{(n)} \cos j\sigma_1\tau, \\
\zeta_n &= \sum_{j=0}^{n+1} B_j^{(n)} \cos j\sigma_1\tau, \qquad (16) \\
\delta_n &= d_n,
\end{aligned}
$$

where $A_j^{(n)}$, $B_j^{(n)}$, d_n are constants.

Case II. Period $2\pi/\sigma_2$.

In this case we take

$$
\begin{aligned}
\rho &= \lambda_2(A_3 \cos \sigma_2\tau + A_4 \sin \sigma_2\tau), \\
\zeta &= \quad A_3 \cos \sigma_2\tau + A_4 \sin \sigma_2\tau,
\end{aligned}
$$

as the generating solutions and proceed as in the previous case. The solutions at the general step are the same as (16) with σ_1 replaced by σ_2.

Case III. Period $n_1(2\pi/\sigma_1) = n_2(2\pi/\sigma_2)$.

In this case, σ_1 and σ_2 must be commensurable, that is

$$\frac{\sigma_1}{\sigma_2} = \frac{n_1}{n_2}$$

where n_1, n_2 are relatively prime integers. Owing to the complexity of the work the construction is omitted. Furthermore, it seems rather hopeless to determine μ so that σ_1 and σ_2 are commensurable.

6. FINAL FORM

When the appropriate substitutions are made in (13) and (6), the solutions take the final form

$$\left.\begin{aligned} r &= r_0\left(1 + \sum_{n=1}^{\infty}\sum_{j=0}^{n} A_j^{(n)} \cos j\sigma\tau\, \epsilon^n\right), \\ z &= \mu r_0\left(1 + \sum_{n=1}^{\infty}\sum_{j=0}^{n} B_j^{(n)} \cos j\sigma\tau\, \epsilon^n\right), \\ \delta &= \sum_{n=2}^{\infty} d_n\, \epsilon^n, \end{aligned}\right\} \tag{17}$$

where $A_j^{(n)}$, $B_j^{(n)}$, d_n are known constants, and σ represents either σ_1 or σ_2.

The initial value r_0 remains arbitrary but c, arising from the integral of areas, together with r_0 must satisfy

$$\frac{c^2}{r_0} = \frac{2(1-\mu^3)}{(1+\mu^2)^{3/2}} = \frac{(4-k^{1/3})^{3/2} - k^2}{4}. \tag{18}$$

Thus r_0 determines the scale of the particular solution. For a given r_0 the initial angular velocity of the pair *one-two* must be chosen so that (18) is satisfied. The parameter ϵ is arbitrary, except for range, and is the scale factor of the initial displacement from the particular solution.

The bodies of pair *one-two* remain diametrically opposite, and pulsate toward and away from the origin as they circulate about it. They move so as to be always equidistant from the origin. Likewise the bodies of pair *three-four* pulsate to and fro in the z-axis at equal distances from the origin.

7. The Convergence of the Solutions

Only the *formal* construction of the solutions was considered in the preceding sections. By the application of McMillan's theorem to the solutions of this type we find that the convergence is assured for sufficiently small numerical values of ϵ_1.

REFERENCES

MacMillan, Trans. Amer. Math. Soc., **13**, no. 2 146-158.

AN ARITHMETICAL IDENTITY

By WALTER H. GAGE

Presented by DANIEL BUCHANAN, F.R.S.C.

1. *Introduction.* In this paper we derive a new arithmetical identity of a type similar to those obtained by Liouville (1), Uspensky (2), Bell (3), and others. This identity involves a function which is restricted by only very general conditions. Hence, by specialization of the function, we can derive from the identity a number of arithmetical theorems.

We shall indicate how to obtain several recursion formulae involving the class-number (4) functions $F(n)$, $G(n)$, $E(n)$. In obtaining these results we make use of formulae for the number of representations (5) of an integer in the forms $xy + yz + zx$, $xy + yz + 2zx$, $xy + 2yz + 2zx$.

2. *Notation.* Throughout this paper n, x, y, z, a, b, c, a', b', c', r, s, t, γ, γ' are integers, of which all but the last two are always positive and n is fixed.

The quadruples (a, b, c, γ), (a', b', c', γ') denote solutions of

$$n = 2ab - c^2 + \gamma^2 \tag{1}$$

$$n = 2a'b' - c'^2 + \gamma'^2 \tag{2}$$

respectively, and the triple (r, s, t) denotes a solution of

$$n = r^2 + s^2 - t^2. \tag{3}$$

3. *The Identity.* Let $\phi(a, b, c, \gamma)$ be a function, finite and uniform for all (a, b, c, γ) and subject only to the conditions

$$\phi(a, b, c, \gamma) = \phi(a, b, -c, -\gamma) = \phi(\gamma, c, b, a). \tag{4}$$

Then

$$\begin{aligned} &\textstyle\sum_1[\phi(a + b - c, c, b - a + \gamma, \gamma) - \phi(b + \gamma, a - c, b - c, a - \gamma)] \\ &= \textstyle\sum_3[\phi(r + s - t, t, 0, s - r) - \phi(s, s - t, r - t, r)] \end{aligned} \tag{5}$$

where $\sum$ with the subscript 1 indicates that ϕ is summed over all solutions (a, b, c, γ) of (1) subject to the restrictions $a > c$, $b > c$, and $\sum$ with the subscript 3 indicates the sum over all solutions (r, s, t) of (3) with $r > t$, $s > t$.

In what follows the restrictions on $\sum$, namely $a > c$, $b > c$, $r > t$, $s > t$, are to be understood. Only additional restrictions will be indicated henceforth.

4. *Derivation.* Noting the one-one correspondence

$$a' = a,\ b' = b + \gamma - c,\quad c' = a - c,\ \gamma' = a - \gamma,$$
$$a = a',\ b = b' + \gamma' - c',\ c = a' - c',\ \gamma = a' - \gamma',$$

under which

$$2ab - c^2 + \gamma^2 = 2a'b' - c'^2 + \gamma'^2$$

we see that

$$\sum_{\substack{1\\ b-a+\gamma>0}} \phi(a + b - c, c, b - a + \gamma, \gamma) = \sum_{\substack{2\\ b'-a'+\gamma'>0}} \phi(a' + b' - c', c', b' - a' + \gamma', \gamma')$$
$$= \sum_{\substack{1\\ b-a+\gamma>0}} \phi(b + \gamma, a - c, b - c, a - \gamma).$$

Now, under the transformation

$$(a, b, c, \gamma) \prec\succ (b, a, c, -\gamma)$$

we have the following correspondences:

$$(a>c,\ b>c>0,\ b-a+\gamma>0) \quad \prec\succ \quad (b>c,\ a>c>0,\ a-b-\gamma>0)$$
$$(a+b-c,\ c,\ b-a+\gamma,\ \gamma) \quad \prec\succ \quad (a+b-c,\ c,\ a-b-\gamma,\ -\gamma)$$
$$(b+\gamma,\ a-c,\ b-c,\ a-\gamma) \quad \prec\succ \quad (a-\gamma,\ b-c,\ a-c,\ b+\gamma)$$

Because of (1) we may therefore write

$$\sum_{\substack{1\\ b-a+\gamma\neq0}} \phi(a + b - c, c, b - a + \gamma, \gamma) = \sum_{\substack{1\\ b-a+\gamma\neq0}} \phi(b + \gamma, a - c, b - c, a - \gamma).$$

Accounting for the terms in each member for which $b - a + \gamma = 0$ we obtain the identity (5).

5. *Simple Applications.* It we make the substitutions

$$a = x + z,\ b = (y + z)/2,\ c = z$$

the left member of (5) is summed over all solutions of $n = xy+yz+zx$ with $y>z$, $y \equiv z \pmod 2$.

Let $\phi(a, b, c, \gamma) = 1$ if $a \equiv b \equiv c \equiv \gamma \equiv 1 \pmod 2$ and equal zero otherwise. After simple reductions (5) we obtain Kronecker's formula (6)

$$F(2m) - 2F(2m - 1^2) + 2F(2m - 2^2) - 2F(2m - 3^2) + \ldots = 0,$$

m being an odd integer greater than zero.

On choosing $\phi(a, b, c, \gamma) = 1$ if $a \equiv \gamma \equiv 0 \pmod 2$, $b \equiv c \equiv 1 \pmod 2$, and equals zero otherwise, we get two formulae. For $m \equiv 3 \pmod 8$ we have

$$F(m) - F(m-1^2) + 2F(m-4^2) - F(m-3^2) + 2F(m-8^2) - F(m-5^2)$$
$$+ 2F(m-12^2) + \ldots = 0,$$

and for $m \equiv 7 \pmod 8$ we have

$$F(m-1^2) - 2F(m-2^2) + F(m-3^2) - 2F(m-6^2) + F(m-5^2)$$
$$- 2F(m-10^2) + \ldots = 0.$$

Finally, the choice $\phi(a, b, c, \gamma) = 1$ if $a \equiv \gamma \equiv 1 \pmod 2$, $b \equiv c \equiv 0 \pmod 2$ and zero otherwise leads to three formulae. For $m \equiv 5 \pmod 8$ the identity yields

$$F(m) - 6E(m-1^2) + 2F(m-2^2) - 6E(m-3^2) + 2F(m-4^2) - \ldots = 0.$$

For $m \equiv 1 \pmod{16}$ it gives

$$F(m) - 3F(m-1^2) + 2F(m-2^2) - 3F(m-3^2) + 2F(m-4^2) - \ldots$$
$$= 6\left[G\left(\frac{m-1^2}{16}\right) + G\left(\frac{m-7^2}{16}\right) + G\left(\frac{m-9^2}{16}\right) + G\left(\frac{m-15^2}{16}\right) + \ldots\right],$$

and for $m \equiv 9 \pmod{16}$ it gives

$$F(m) - 3F(m-1^2) + 2F(m-2^2) - 3F(m-3^2) + 2F(m-4^2) - \ldots$$
$$= 6\left[G\left(\frac{m-3^2}{16}\right) + G\left(\frac{m-5^2}{16}\right) + G\left(\frac{m-11^2}{16}\right) + G\left(\frac{m-13^2}{16}\right) + \ldots\right]$$

The University of British Columbia.

REFERENCES

(1) **Liouville.** Journal de Mathématiques, 1858-65.

(2) **Uspensky** and **Heaslet.** Elementary Number Theory, chapter XIII, McGraw-Hill. 1939.

(3) **Bell, E. T.** Algebraic arithmetic. Amer. Math. Soc. Colloquium Pub., **7**: 89-106. 1927.

(4) **Smith, H. J. S.** Mathematical Papers, **1**.

(5) **Bell, E. T.** Numbers of representations of integers in a certain triad of ternary quadratic forms. Trans. Amer. Math. Soc., **32**, no 1: 1-5. 1930. See also **Gage**, An arithmetical identity for the form $ab - c^2$. Bull. Amer. Math. Soc., **48**, no. 12: 898-900. 1942.

(6) **Dickson.** History of the Theory of Numbers, **3**, chapter VI, p. 108, formula III.

A SHORT TABLE OF THE TORONTO FUNCTION

By A. H. HEATLEY

Communicated by D. McINTOSH, F.R.S.C.

1. INTRODUCTION

SIX-FIGURE tables of the confluent hypergeometric function

$$M(a, \gamma, x) = 1 + \frac{a}{\gamma}\frac{x}{1!} + \frac{a(a+1)}{\gamma(\gamma+1)}\frac{x^2}{2!} + \dots \tag{1}$$

have been published by Airey (1) for the range $a = -4\left(\tfrac{1}{2}\right)4$,* with $\gamma = -\tfrac{3}{2}, -\tfrac{1}{2}, \tfrac{1}{2}, 1, \tfrac{3}{2}, 2, 3, 4$, and $x = 0.00\ (0.02)\ 0.10\ (0.05)\ 1.0\ (0.1)\ 2.0\ (0.2)\ 3.0\ (0.5)\ 8.0$. He discussed various differential equations solvable in terms of the M function. A seven-figure table of the function

$$H(m, a, x) = e^{-ix} M(m + 1 - ia, 2m + 2, 2ix) \qquad (2)\dagger$$

has been published by Lowan and Horenstein (8A) for the range $m = 0(1)3, a = 0(1)10, x = 0(1)10$. This function occurs in problems of molecular physics. They also list a number of other problems which lead to the M function.

The function tabulated in the present paper, which is related to the M function, has its origin as a generalization, ($T(m, n, p, a)$, see Eq. (3)) of a Bessel function integral encountered in a study of the theory of collectors in electrical discharges (5, Part II). The original function is here transformed to ($T(m, n, r)$ (see Eq. (6)) which is so defined that it is asymptotic to unity for large values of r; this is a distinct advantage in tabulation and interpolation. Differential equations are given for both the original and modified functions; the relations of the modified function to other tabulated functions, its recurrence and asymptotic formulae are also shown. A short exploratory table of nearly one thousand five- or six-figure entries of the T function is given, preceded by a table of about one hundred M functions required for the computation of T.

The calculation of the numerical value of an M function from the corresponding $T(m, n, r)$ or *vice versa*, is sufficiently tedious that the tables should not be considered as overlapping (if and when they are both available), and is, however, sufficiently easy (for the factors are well tabulated functions) that the tables may be considered as supplementary; but the fact that the T function is, in general, more easily interpolable gives it considerable practical advantage. Another justification for the tabulation of $T(m, n, r)$ is the group of differential equations and integrals easily solvable in terms of that function.

---

*The number in brackets indicates the interval between successive values of the parameter.

†The authors used the symbol $F(a, \gamma\ ;\ u)$ instead of M.

The original T function (5) was defined, and evaluated in terms of the M function, thus:

$$T(m, n, p, a) \equiv \int_0^\infty t^{m-n} e^{-p^2t^2} I_n(2at)\,dt, \quad (T1)^{*\dagger} \qquad (3)$$

$$= \frac{\Gamma\left(\frac{m+1}{2}\right)(a/p)^n}{\Gamma(n+1)\,2p^{m-n+1}} M\left(\frac{m+1}{2}, n+1, a^2/p^2\right); m > -1. \quad (T55) \qquad (4)$$

From Eq. (4) it is easily seen that

$$T(m, n, 1, r) = p^{m-n+1} T(m, n, p, a), \qquad (5)$$

where $r = a/p$. Since this is a very simple relation, it is sufficient to tabulate a three-parameter function. It is also obvious that the function need be tabulated for positive values only of r.

2. THE NEW FUNCTION

For convenience of interpolation and tabulation the function to be tabulated will be defined by

$$T(m, n, r) \equiv 2r^{n-m+1} e^{-r^2} T(m, n, 1, r), \qquad (6)^\ddagger$$

$$= r^{2n-m+1} e^{-r^2} \Gamma\left(\frac{m+1}{2}\right) M\left(\frac{m+1}{2}, n+1, r^2\right) \div \Gamma(n+1). \qquad (7)$$

Anticipating the result in Eq. (42) we have

$$T(m, n, r) \doteq 1 \text{ (for large } r). \qquad (8)$$

The expressions in Eqs. (4) and (7) are indeterminate when n is a negative integer; in such cases the parameters are changed by the relation

$$T(m, n, r) = T(m - 2n, -n, r). \qquad (T2) \qquad (9)$$

If Eq. (7) is taken as the definition of T, it is easily shown that m

---

[*] In the former paper (5), the abbreviated symbol $T(m, n)$ was used. Equations from Part I of that paper, or simple transformations thereof, are indicated by the prefix T as above.

[†] Watson (12, Art. 13.3) studied the integral analogous to that on the right of Eq. (3) with the Bessel function J instead of I. He gave its value for several special cases, and references to earlier work.

[‡] "The Toronto Function" is suggested as the name of the function on the left of Eq. (6).

may take any value except the odd negative integers, at which points the function becomes infinite.

3. DIFFERENTIAL EQUATIONS

One solution of the differential equations

$$y'' + \left(\frac{1}{a} - \frac{2a}{p^2}\right)y' - \left(\frac{2m-2n+2}{p^2} + \frac{n^2}{a^2}\right)y = 0, \qquad (T11) \quad (10)$$

$$y'' + \left(\frac{1}{a} + \frac{2a}{p^2}\right)y' - \left(\frac{2m-2n-2}{p^2} + \frac{n^2}{a^2}\right)y = 0, \qquad (T16) \quad (11)$$

$$y'' + \left(\frac{1}{a}\right)y' - \left(\frac{2m-2n+a^2/p^2}{p^2} + \frac{n^2}{a^2}\right)y = 0, \qquad (T21) \quad (12)$$

is given by (5)

$$y = T(m, n, p, a),\ e^{-a^2/p^2}T(m, n, p, a), \text{ and } e^{-a^2/2p^2}T(m, n, p, a),$$

respectively, where $y'' = \partial^2 y/\partial a^2$ and $y' = \partial y/\partial a$.

The partial derivatives of $T(m, n, r)$ with respect to m and to n seem to be intractable, but it is easily shown that

$$\frac{\partial}{\partial r}T(m,n,r) = 2r\,T(m+2, n+1, r) - \left(\frac{2r^2+m-2n-1}{r}\right)T(m,n,r), \qquad (13a)$$

$$= \frac{m-2n-1}{r}\left\{T(m, n+1, r) - T(m, n, r)\right\}, \qquad (13b)$$

$$= 2r\left\{T(m, n-1, r) - T(m, n, r)\right\} - (m-1)\,T(m, n, r), \qquad (13c)$$

$$= (2n-m+1)\,r^{2n-m}\,\Gamma\left(\frac{m+1}{2}\right) \div \Gamma(n+1) \text{ when } r \to 0. \qquad (13d)$$

$$\frac{\partial^2}{\partial r^2}T(m, n, r) - \left(\frac{2n-2m+1-2r^2}{r}\right)\frac{\partial}{\partial r}T(m,n,r) + \frac{(m-1)(m-2n-1)}{r^2}T(m, n, r) = 0. \qquad (14)$$

4. RELATION TO OTHER FUNCTIONS

Eq. (7) expresses T in terms of the M function. Although tables of the latter function have been published by Airey (1), his values provide only a very limited number of those required for the present computations. However, for many special cases the T function may be expressed in terms of other tabulated functions. The first, second, and fifth cases below are readily derived from known relations.

Case I. The exponential function

$$T(2n+1, n, 1, r) = \tfrac{1}{2} r^n e^{r^2}. \qquad (T28) \qquad (15)$$

$$T(2n+1, n, r) = 1. \qquad (16)$$

Combining Eq. (16) with Eq. (39) it is easily shown that

$$T(2n+1, n-1, r) = 1 + n/r^2. \qquad (17)$$

The special formulae given below are obtained by putting $n_0 = 0$ in Eqs. (41a) and (41b).

$$T(1, n, r) = 1 - e^{-r^2} \sum_{s=0}^{n-1} r^{2s}/s!; \; n = 1, 2, 3, \ldots. \qquad (18)$$

$$T(1, n, r) = 1; \; n = 0, -1, -2, \ldots. \qquad (19)$$

Case II. The Bessel function $I_n(x)$

$$T(m, m, 1, r) = \tfrac{1}{2} \pi^{\frac{1}{2}} e^{r^2/2} I_{m/2}(r^2/2). \qquad (T34) \qquad (20)$$

$$T(m, m, r) = r \pi^{\frac{1}{2}} e^{-r^2/2} I_{m/2}(r^2/2). \qquad (21)$$

$$T(m, m+1, 1, r) = \frac{r\pi^{\frac{1}{2}}}{2(m+1)} e^{r^2/2} \left\{ I_{m/2}(r^2/2) - I_{m/2+1}(r^2/2) \right\}. \qquad (T45) \qquad (22)$$

$$T(m, m+1, r) = \frac{r^3 \pi^{\frac{1}{2}}}{m+1} e^{-r^2/2} \left\{ I_{m/2}(r^2/2) - I_{m/2+1}(r^2/2) \right\}, \qquad (23a)$$

$$= \frac{r^3 \pi^{\frac{1}{2}}}{m+1} e^{-r^2/2} \left\{ -I_{m/2-1}(r^2/2) + \left(\frac{2m}{r^2} + 1 \right) I_{m/2}(r^2/2) \right\}. \qquad (23b)$$

Case III. The error function

$$T(1, \tfrac{1}{2}, 1, r) = \int_0^\infty t^{1/2} e^{-t^2} I_{1/2}(2rt)\, dt,$$

$$= \tfrac{1}{2} \int_0^\infty t^{1/2} e^{-t^2} (e^{2rt} - e^{-2rt}) / (\pi r t)^{\frac{1}{2}}.\, dt,$$

$$= (e^{r^2}/2r^{\frac{1}{2}}) \operatorname{Erf}(r), \qquad (24)$$

where

$$\operatorname{Erf}(r) = \frac{2}{\pi^{\frac{1}{2}}} \int_0^r e^{-x^2} dx. \qquad (25)$$

Whence

$$T(1, \tfrac{1}{2}, r) = \operatorname{Erf}(r). \qquad (26)$$

Combining this result with Eqs. (41a) and (41b) in which we put $n_0 = \frac{1}{2}$, it may be shown that

$$T(1, n, r) = \operatorname{Erf}(r) - e^{-r^2} \sum_{s=1/2}^{n-1} r^{2s}/s!; \; n = \tfrac{3}{2}, \tfrac{5}{2}, \ldots. \qquad (27)$$

$$T(1, n, r) = \operatorname{Erf}(r) + e^{-r^2} \sum^{-1/2} r^{2s}/s!; \; n = -\tfrac{1}{2}, -\tfrac{3}{2}, \ldots. \qquad (28)$$

Case IV. The F function of Miller and Gordon (9)

$$M\left(\tfrac{1}{2}, \tfrac{3}{2}, r^2\right) = 1 + \frac{1}{3}r^2 + \frac{1}{5 \cdot 2!}r^4 + \frac{1}{7 \cdot 3!}r^6 + \ldots,$$

$$= \frac{1}{r}\int_0^r e^{r^2}\,dr = \frac{1}{r}\,e^{r^2}F(r). \tag{29}$$

Whence, by Eq. (7),

$$T\left(0, \tfrac{1}{2}, r\right) = 2r^2 e^{-r^2} M\left(\tfrac{1}{2}, \tfrac{3}{2}, r^2\right) = 2r\,F(r). \tag{30}$$

Case V. The incomplete Gamma function

$$T(1, n, r) = 2r^n e^{-r^2} T(1, n, 1, r) = I(r,^2\, n - 1), \quad (T78) \tag{31}$$

where

$$I(c, n-1) = \Gamma_c(n)/\Gamma(n) = \int_0^c t^{n-1}e^{-t}\,dt \div \int_0^\infty t^{n-1}e^{-t}\,dt. \tag{32}$$

Unfortunately, the only available table of the I Gamma function (11) was tabulated for convenience of use in a special way and is not convenient as a source of the related T function. It is hoped that for some purposes the present limited table of T will be found useful by those requiring a table of the I Gamma function, although for this purpose there is the handicap* that r^2 in the I function corresponds to r in the T function.

5. RECURRENCE FORMULAE

Two basic recurrence formulae for the original T function have been given (5). By changing these to the $T(m, n, r)$ form and then appropriately manipulating them, six other useful formulae are easily derived.

$$T(m+2, n+2, r) = T(m, n, r) - \frac{n+1}{r^2}T(m, n+1, r). \quad (T37) \tag{33}$$

$$T(m+2, n+1, r) = T(m+2, n, r) - \frac{m+1}{2r^2}T(m, n, r). \quad (T41) \tag{34}$$

$$T(m+2, n+1, r) = \frac{m-2n-1}{2r^2}T(m, n+1, r) + T(m, n, r). \tag{35}$$

$$T(m+2, n, r) = \frac{m-2n-1}{2r^2}T(m, n+1, r) + \frac{m+1+2r^2}{2r^2}T(m, n, r). \tag{36}$$

---

*Of course, $T(m, n, r)$ could be tabulated against the argument $r^{\frac{1}{2}}$ instead of r; but this would cause some loss of convenience elsewhere. The computer who may some day plan to extend the present table should carefully consider the case pro and con.

$$T(m+2, n+1, r) = \frac{2r^2}{m+1+2r^2} T(m+2, n, r) + \frac{(m+1)(m-2n-1)}{2r^2(m+1+2r^2)} T(m, n+1, r). \qquad (37)$$

$$T(m+1, n, r) = \frac{m-n+2+r^2}{r^2} T(m+2, n, r) - \frac{(m+1)(m-2n+1)}{4r^4} T(m, n, r). \qquad (38)$$

$$T(m, n, r) = \frac{r^2+n+1}{r^2} T(m, n+1, r) + \frac{m-2n-3}{2r^2} T(m, n+2, r). \qquad (39)$$

$$T(m, n, r) = \frac{2(n+1-r^2)}{m+1} T(m+2, n+1, r) + \frac{2r^2}{m+1} T(m+1, n+2, r). \qquad (40)$$

Eqs. (34) to (37) give the relations between the functions at the corners of the rectangle (m, n), $(m+2, n)$, $(m+2, n+1)$, $(m, n+1)$ in groups of three. Eqs. (38) to (40) are the "m," "n," and "diagonal" recurrence formulae, respectively.

A special recurrence formula is also given.

$$T(1, n+1, r) = T(1, n, r) - e^{-r^2} r^{2n}/n!. \qquad (T'46) \quad (41)$$

From this it is easily shown that

$$T(1, n, r) = T(1, n_0, r) - e^{-r^2} \sum_{s=n_0}^{n-1} r^{2s}/s!\,; \; n > n_0, \qquad (41a)$$

$$T(1, n, r) = T(1, n_0, r) + e^{-r^2} \sum_{s=n}^{n_0-1} r^{2s}/s!\,; \; n < n_0. \qquad (41b)$$

where n_0 is a number differing from n by an integer.

6. VALUES OF THE FUNCTION FOR SPECIFIED VALUES OF m AND n.

Combining the data in Sections 4 and 5 the values of T have been expressed in terms of other functions for $m = -\frac{1}{2}, 0, \frac{1}{2}, 1$ and $n = 0, \pm\frac{1}{2}, \pm 1, \pm\frac{3}{2}, \pm 2$, with the exception of ten functions which remain in terms of M. The relations are given in Table I.

7. SEMI-CONVERGENT SERIES

The asymptotic expansion for the M function has been given by Airey (1, p. 276, 1926). Combining this result with the value of T in Eq. (7) we have

TABLE IA

T (m, n, r) IN TERMS OF OTHER FUNCTIONS

$\rho = r^2/2 \qquad I_n = I_n(r^2/2)$

n	m = $-\frac{1}{2}$	0
2	$2\Gamma(\frac{5}{4}) r^{\frac{11}{2}} e^{-2\rho} M(\frac{1}{4}, 3, r^2)$	$\frac{2}{3} r^5 \pi^{\frac{1}{2}} e^{-\rho} \{I_0 - (1+\frac{1}{2\rho}) I_1\}$
$\frac{3}{2}$	$\frac{8}{5} r^5 \pi^{\frac{1}{2}} e^{-\rho} \{I_{-\frac{1}{4}} - (1+\frac{1}{4\rho}) I_{\frac{3}{4}}\}$	$2r(r^2+\frac{1}{2})F(r) - r^2$
1	$4\Gamma(\frac{5}{4}) r^{\frac{7}{2}} e^{-2\rho} M(\frac{1}{4}, 2, r^2)$	$r^3 \pi^{\frac{1}{2}} e^{-\rho} \{I_0 - I_1\}$
$\frac{1}{2}$	$2r^3 \pi^{\frac{1}{2}} e^{-\rho} \{I_{-\frac{1}{4}} - I_{\frac{3}{4}}\}$	$2r\,F(r)$
0	$4\Gamma(\frac{5}{4}) r^{\frac{3}{2}} e^{-2\rho} M(\frac{1}{4}, 1, r^2)$	$r \pi^{\frac{1}{2}} e^{-\rho} I_0$
$-\frac{1}{2}$	$r \pi^{\frac{1}{2}} e^{-\rho} I_{-\frac{1}{4}}$	One
-1	$\Gamma(\frac{5}{4}) r^{\frac{3}{2}} e^{-2\rho} M(\frac{5}{4}, 2, r^2)$ $T(\frac{3}{2}, 1, r)$	$\frac{1}{2} r \pi^{\frac{1}{2}} e^{-\rho} \{I_0 + I_1\}$ $T(2, 1, r)$
$-\frac{3}{2}$	$\frac{1}{2} r \pi^{\frac{1}{2}} e^{-\rho} \{(1-\frac{1}{2\rho}) I_{-\frac{1}{4}} + I_{\frac{3}{4}}\}$	$1 - \frac{1}{4\rho}$
-2	$\frac{5}{8}\Gamma(\frac{5}{4}) r^{\frac{3}{2}} e^{-2\rho} M(\frac{9}{4}, 3, r^2)$ $T(\frac{7}{2}, 2, r)$	$\frac{1}{2} r \pi^{\frac{1}{2}} e^{-\rho} \{I_0 + (1-\frac{1}{2\rho}) I_1\}$ $T(4, 2, r)$

TABLE IB

T (*m*, *n*, *r*) IN TERMS OF OTHER FUNCTIONS

$\rho = r^2/2 \qquad I_n = I_n(r^2/2)$

n	m = $\frac{1}{2}$	1
2	$\frac{2}{3}\Gamma\left(\frac{7}{4}\right) r^{\frac{9}{2}} e^{-2\rho} M\left(\frac{3}{4}, 3, r^2\right)$	$1-(1+r^2)e^{-2\rho}$
$\frac{3}{2}$	$\frac{2}{3} r^3 \pi^{\frac{1}{2}} e^{-\rho}\left\{\left(1+\frac{1}{2\rho}\right) I_{\frac{1}{4}} - I_{-\frac{3}{4}}\right\}$	$\mathrm{Erf}(r) - 2r\pi^{-\frac{1}{2}} e^{-2\rho}$
1	$\frac{4}{3}\Gamma\left(\frac{7}{4}\right) r^{\frac{5}{2}} e^{-2\rho} M\left(\frac{3}{4}, 2, r^2\right)$	$1-e^{-2\rho}$
$\frac{1}{2}$	$r\pi^{\frac{1}{2}} e^{-\rho} I_{\frac{1}{4}}$	$\mathrm{Erf}(r)$
0	$\frac{4}{3}\Gamma\left(\frac{7}{4}\right) r^{\frac{1}{2}} e^{-2\rho} M\left(\frac{3}{4}, 1, r^2\right)$	One
$-\frac{1}{2}$	$\frac{1}{2} r \pi^{\frac{1}{2}} e^{-\rho}\left\{I_{\frac{1}{4}} + I_{-\frac{3}{4}}\right\}$	$\mathrm{Erf}(r) + r^{-1}\pi^{-\frac{1}{2}} e^{-2\rho}$
-1	$\Gamma\left(\frac{7}{4}\right) r^{\frac{1}{2}} e^{-2\rho} M\left(\frac{7}{4}, 2, r^2\right)$ $T\left(\frac{5}{2}, 1, r\right)$	One $T(3, 1, r)$
$-\frac{3}{2}$	$\frac{1}{2} r \pi^{\frac{1}{2}} e^{-\rho}\left\{I_{\frac{1}{4}} + \left(1-\frac{1}{4\rho}\right) I_{-\frac{3}{4}}\right\}$	$\mathrm{Erf}(r) + r^{-1}\pi^{-\frac{1}{2}}\left(1-\frac{1}{4\rho}\right) e^{-2\rho}$
-2	$\frac{7}{8}\Gamma\left(\frac{7}{4}\right) r^{\frac{1}{2}} e^{-2\rho} M\left(\frac{11}{4}, 3, r^2\right)$ $T\left(\frac{9}{2}, 2, r\right)$	One $T(5, 1, r)$

$$T(m, n, r) \approx 1 + \frac{(1-m)(2n+1-m)}{1!\,4r^2} + \frac{(1-m)(3-m)(2n+1-m)(2n+3-m)}{2!\,(4r^2)^2} + \dots$$
$$+ \frac{e^{-r^2}\,r^{2n-2m}}{i^{m+1}} \frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{2n-m+1}{2}\right)} \left\{1 - \frac{(m+1)(m-2n+1)}{1!\,4r^2} + \frac{(m+1)(m+3)(m-2n+1)(m-2n+3)}{2!\,(4r^2)^2} - + \dots\right\}. \quad (42)$$

For the four values of m to be tabulated this reduces to

$$T(-\tfrac{1}{2}, n, r) = 1 + \frac{3\,(4n+3)}{1!\,16r^2} + \frac{3 \cdot 7\,(4n+3)(4n+7)}{2!\,(16r^2)^2} + \dots\dots \quad (43)$$

$$T(0, n, r) = 1 + \frac{1\,(2n+1)}{1!\,4r^2} + \frac{1 \cdot 3\,(2n+1)(2n+3)}{2!\,(4r^2)^2} + \dots\dots \quad (44)$$

$$T(\tfrac{1}{2}, n, r) = 1 + \frac{1\,(4n+1)}{1!\,16r^2} + \frac{1 \cdot 5\,(4n+1)(4n+5)}{2!\,(16r^2)^2} + \dots\dots \quad (45)$$

$$T(1, n, r) = 1 - \frac{e^{-r^2}\,r^{2n-2}}{\Gamma(n)} \left\{1 - \frac{(1-n)}{r^2} + \frac{(1-n)(2-n)}{r^4} - + \dots\dots\right\}. \quad (46)$$

The usefulness of semi-convergent series for computation is, in general, greatly increased by Airey's "converging factor" (C.F.) (2). Slight modification of series for which C.F.'s are already given (1, pp. 273-97, 1926) provides the C.F. for $2rF(r) = T(0, \frac{1}{2}, r)$, but existing tables (9, 6) render it unnecessary. The C.F. for $I(r^2, n-1) = T(1, n, r)$ is also given (2),

$$T(1, n, r) = 1 - \frac{e^{-r^2}}{r\pi^{\frac{1}{2}}} \left[1 - \frac{1-n}{r^2} + \dots (-1)^{s-1} \frac{(1-n)(2-n)\dots(s-1-n)}{r^{2s-2}}\,\text{C.F.}\right]. \quad (47)$$

$$\text{C.F.} = \frac{1}{2} + \frac{1}{4k}\left(\frac{1}{2} + h\right) - \frac{1}{8k^2}\left(\frac{1}{4} + \frac{h}{2} + h^2\right) - \dots$$

where $k = s - n$, $h = r^2 - k$, and sth term is the least of the semi-convergent series.

The C.F.'s for other T functions or related functions do not seem to be available.

8. COMPUTATION OF THE TABLE

The computation was done on an eight-column Monroe hand machine, supplemented when necessary by hand work; the work was carried to eight decimal places. The tables used as a basis for the calculations are shown in Table II.

In most cases, for values of r greater than four, the semi-convergent series of Eqs. (43) to (45) were used instead of the expressions in Table I.

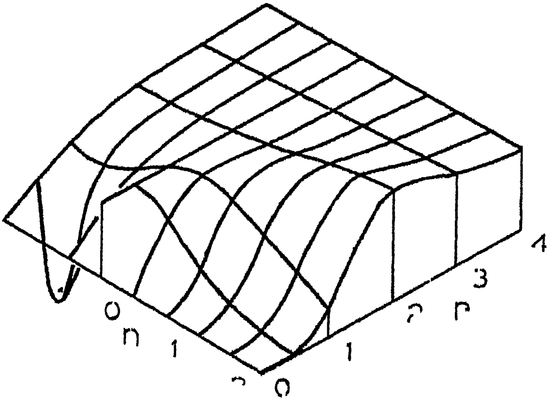


FIGURE 1. The Graph of $T(0, n, r)$. (The line for $r = 0.5$ is omitted between $n = 0.5$ and 1.15.)

The computed values were checked before rounding off by the use of the recurrence formula, Eq. (37). For values of r greater than four, when this method becomes less suitable because of the large multipliers, the values were checked by successive differences for variable n and constant m and r. The Tables at the end of this paper give these values, rounded off to five decimal places. The printer's proof sheets were compared with the table to eight decimals which was used for checking rather than against the rounded off table.

A graph of the function $T(0, n, r)$ is shown in Fig. 1. Owing to the wide variations of $T(0, n, 0)$, its line is not shown, but instead $T(0, n, \frac{1}{2})$ is shown. The approximate values of this function for part of the range are shown in Table III.

9. APPENDIX. A SHORT TABLE OF THE M FUNCTION

Ten groups of the M function are required to complete the computations of $T(m, n, r)$ for the range of m and n shown in Table I. The computations were continued only as far as $r = 2$ which required 19 terms of the series given in Eq. (1). The gap between $r = 2$ and the lowest values of r computable to the required accuracy by the semi-convergent series is left for someone with better facilities than the present author.

TABLE II

SOURCE TABLES FOR COMPUTATIONS

Function	*Author*	*Reference*
Exponential Function	Lowan	7
Error Function	Lowan	8
Bessel Functions I_0, I_1	BAAS	4, 12
Bessel Functions $I_{\pm\frac{1}{4}}$, $I_{\pm\frac{3}{4}}$	Lowan	Private Communication
The F Function	Miller & Gordon	9, 6
Square Roots	Milne-Thomson	10
Gamma (5/4) & (7/4)	BAAS	3, page 40
The M Function	Computed.	See Appendix

TABLE III

APPROXIMATE VALUES OF $T(0, n, \frac{1}{2})$

n	$T(0, n, \frac{1}{2})$	n	$T(0, n, \frac{1}{2})$	n	$T(0, n, \frac{1}{2})$
-0.5	+1.00	1.0	+0.42	−1.5	−1.00
0.6	0 97	1.1	+0.16	1.6	−1.16
0.7	0.90	1 2	−0.13	1.7	−1.17
0.8	0.80	1.3	−0.45	1.8	−0.98
0.9	0.63	1.4	−0.75	1.9	−0.50
-1.0	+0.42	1.5	−1.00	−2.0	+0.32

The work of computation was somewhat reduced by noting the relation,

$$M(a, \gamma, x) = e^{x} M(\gamma - a, \gamma, -x). \tag{48}$$

Thus the same three series served for $M(\frac{1}{4}, 3, x)$ and $M(\frac{11}{4}, 3, x)$ for $M(\frac{1}{4}, 2, x)$ and $M(\frac{7}{4}, 2, x)$, for $M(\frac{3}{4}, 3, x)$ and $M(\frac{9}{4}, 3, x)$, respectively, by adding or subtracting the sums of the odd and even terms of the series. From the values of $M(\frac{1}{4}, 3, x)$ and $M(\frac{1}{4}, 2, x)$ the values of $M(\frac{1}{4}, 1, x)$, $M(\frac{5}{4}, 2, x)$, and $M(\frac{9}{4}, 3, x)$ were computed by the recurrence formulae:—

$$M(\tfrac{1}{4}, 1, x) = (r^2 + 1) M(\tfrac{1}{4}, 2, x) - 0.875\, r^2 M(\tfrac{1}{4}, 3, x)$$

$$M(\tfrac{5}{4}, 2, x) = (4r^2 + 1) M(\tfrac{1}{4}, 2, x) - 3.5\, r^2 M(\tfrac{1}{4}, 3, x)$$

$$M(\tfrac{9}{4}, 3, x) = 1.6\,(4r^2 - 1) M(\tfrac{1}{4}, 2, x) - 1.4\,(4r^2 - 3) M(\tfrac{1}{4}, 3, x).$$

A comparison of the values of $M\left(\frac{9}{4}, 3, x\right)$ computed by the series and by the recurrence formulae provided a check on the work. Similarly, using variations of the above recurrence formulae, the five other required values of M were computed and checked. This method was not applicable above $r = 2$ (and $x = 4$) for its accuracy falls off with increasing r. Even at $r = 2$, there is an uncertainty of a few units in the seventh place of the value of $M\left(\frac{9}{4}, 3, x\right)$ and $M\left(\frac{7}{4}, 3, x\right)$ calculated from these formulae. The values were computed to eight decimal places and all digits were used in the T computations. Table IV gives them as rounded off to five places.

TABLE IV

THE CONFLUENT HYPERGEOMETRIC FUNCTION

$M(a, \gamma, x)$

x	$a, \gamma =$	$\frac{1}{4}, 3$	$\frac{1}{4}, 2$	$\frac{1}{4}, 1$	$\frac{5}{4}, 2$	$\frac{9}{4}, 3$
0.00		1.00000	1.00000	1.00000	1.00000	1.00000
0.04		1.00335	1.00504	1.01013	1.02538	1.03019
0.16		1.01367	1.02069	1.04208	1.10627	1.12817
0.36		1.03178	1.04862	1.10111	1.25858	1.31385
0.64		1.05923	1.09209	1.19787	1.51518	1.63100
1.00		1.09861	1.15685	1.35241	1.93910	2.16386
1.44		1.15120	1.25284	1.60262	2.65198	3.07722
1.96		1.23314	1.39778	2.02260	3.80705	4.70512
2.56		1.34704	1.62449	2.76446	6.18439	7.75757
3.24		1.51962	1.99685	1.15851	10.64350	13.82711
4.00		1.79041	2.64661	6.96661	19.92660	26.69856

$e^{-x}M(a, \gamma, x)$

x	$a, \gamma =$	$\frac{3}{4}, 3$	$\frac{3}{4}, 2$	$\frac{3}{4}, 1$	$\frac{7}{4}, 2$	$\frac{11}{4}, 3$
0.00		1.00000	1.00000	1.00000	1.00000	1.00000
0.04		0.97048	0.97537	0.99012	0.99504	0.99660
0.16		0.88746	0.90575	0.96192	0.98065	0.98699
0.36		0.76576	0.80263	0.91928	0.95810	0.97160
0.64		0.62501	0.68139	0.86748	0.92951	0.95153
1.00		0.48432	0.55734	0.81197	0.89685	0.92797
1.44		0.35815	0.44248	0.75730	0.86224	0.90213
1.96		0.25448	0.34402	0.70654	0.82739	0.87512
2.56		0.17525	0.26452	0.66127	0.79353	0.84785
3.24		0.11819	0.20311	0.62187	0.76145	0.82099
4.00		0.07895	0.15707	0.58794	0.73157	0.79501

REFERENCES

(1) AIREY, J. R. British Association for the Advancement of Science, Annual reports. 1926, 1927.

(2) — — — — — — Phil. Mag., 24:521-552. 1937.

(3) BRITISH ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE, Mathematical Tables. Vol. 1. London. 1931.

(4) — — — — — — — — — — — — — — — — Mathematical Tables. Vol. 6, Part I. London, Cambridge University Press. 1937.

(5) HEATLEY, A. H. University of Toronto Studies, Mathematical Series No. 7. Toronto. 1939.

(6) — — — — — — Can. Jour. Research, 18A:123-127. 1940.

(7) LOWAN, A. N. Federal Works Agency, Works Projects Administration, Tables of the Exponential Function e^x. New York. 1939.

(8) — — — — — — Federal Works Agency, Works Projects Administration, Tables of Probability Functions. Vol. I. New York. 1941.

(8A) — — — — — — and HORENSTEIN, WILLIAM. Jour. Math. and Phys., 21:264-283. 1942.

(9) MILLER, W. L. and GORDON, A. R. Jour. Phys. Chem., 35:2785-2884. 1931.

(10) MILNE-THOMSON, L. M. Standard Table of Square Roots London, G. Bell and Sons, Ltd. 1929.

(11) PEARSON, R. Tables of the Incomplete Gamma Function. London, Cambridge University Press. 1922.

(12) WATSON, G. N. A Treatise on the Theory of Bessel Functions. London, Cambridge University Press. 1922.

THE TORONTO FUNCTION $T(-\frac{1}{2}, n, r)$

r	$n = -2$	$-\frac{3}{2}$	-1	$-\frac{1}{2}$	0	$\frac{1}{2}$	1	$\frac{3}{2}$	2
0 0	Zero	$-\infty$	Zero	Zero	Zero	Zero	Zero	Zero	Zero
2	0 05017	−10 75575	0 97987	0 89679	0 31472	0 07079	0 01253	0 00188	0 00025
4	0 13778	− 3 10970	0 21626	1 19679	0 81449	0 36257	0 12764	0 03825	0 0 014
6	0 24134	− 1 10936	0 36990	1 33778	1 29447	0 84833	0 44879	0 19857	0 07860
8	0 34861	− 0 24271	0 [illegible]817	1 37423	1 63861	1 39090	0 95611	0 56489	0 20675
1 0	0 45096	−0 21425	0 6[illegible]39	1 34593	1 80383	1 83486	1 51294	1 12[illegible]	0 75267
2	0 54293	0 47813	0 74895	1 28626	1 80967	2 08233	2 03716	1 75090	[illegible]5128
4	0 62194	0 63796	0 82430	1 21970	1 71106	2 12144	2 31767	2 26249	2 40578
6	0 68756	0 73796	0 87700	1 16061	1 56810	2 00674	2 35896	2 535[illegible]	2 [illegible]0188
8	0 74085	0 80246	0 9[illegible]243	1 11468	1 42598	1 81700	2 21853	2 51720	2 73509
2 0	0 78353	0 84554	0 93567	1 08198	1 30849	1 61900	1 98838	2 36608	2 69021
2		0 87546		1 05993		1 46140		2 09653	
4		0 89707		1 04539		1 32649		1 82591	
6		0 91327		1 03575		1 24046		1 60325	
8		0 92779		1 02916		1 18342		1 44090	
3 0		0 93570		1 02144		1 14561		1 33068	
2		0 94372		1 02091		1 11978		1 25788	
4		0 95030		1 01816		1 10128		1 20911	
6		0 95574		1 01596		1 08736		1 17529	
8		0 96040		1 01415		1 07647		1 15012	
4		0 96[illegible]2		1 01265		1 06767		1 13130	
5	0 96284	0 97724	0 93229	1 00786	1 02404	1 04088	1 05841	1 07649	1 09577
6	0 97412	0 98428	0 99469	1 00538	1 01638	1 02760	1 03916	1 0510	1 06326
10	0 99065	0 99436	0 99811	1 00190	1 00571	1 00967	1 01535	1 01757	1 02133
25	0 99850	0 99946	0 99970	1 00030	1 00090	1 00156	1 00211	1 0027	1 00332
50	0 99963	0 99977	0 99992	1 00008	1 00023	1 00038	1 00053	1 00068	00083

THE TORONTO FUNCTION $T(0, n, r)$

r	$n = -2$	$-\frac{3}{2}$	-1	$-\frac{1}{2}$	0	$\frac{1}{2}$	1	$\frac{3}{2}$	2
0 0	Zero	$-\infty$	Zero	One	Zero	Zero	Zero	Zero	Zero
2	0 13205	−11 50000	0 17549	for	0 34751	0 07790	0 01376	0 00207	0 00027
4	0 25899	− 2 12500	0 34086	all	0 65552	0 28795	0 10069	0 03005	0 00795
6	0 37640	− 0 38389	0 48788	values	0 89550	0 56972	0 29348	0 12996	0 05117
8	0 48116	+ 0 21875	0 61152	of r	1 05618	0 85136	0 56916	0 33055	0 17165
1 0	0 57165	0 50000	0 71027		1 14330	1 07616	0 86605	0 61424	0 39253
2	0 64765	0 65278	0 78563		1 17388	1 21746	1 11815	0 92187	0 69194
4	0 71007	0 74490	0 84103		1 16870	1 27822	1 28447	1 18442	1 00758
6	0 76051	0 80469	0 88068		1 14608	1 27981	1 35882	1 35621	1 26897
8	0 80035	0 84568	0 90815		1 11874	1 24838	1 36134	1 42896	1 43158
2 0	0 83298	0 87500	0 92837		1 09363	1 20536	1 32209	1 42413	1 49063
2	0 85860	0 89669	0 94244		1 07338	1 16385	1 26759	1 37494	1 47169
4	0 87913	0 91319	0 95268		1 05812	1 12950	1 21464	1 31069	1 41082
6	0 89571	0 92604	0 96034		1 04693	1 10326	1 17071	1 24966	1 33830
8	0 90922	0 93622	0 96621		1 03873	1 08388	1 13706	1 19959	1 27197
3 0	0 92033	0 94444	0 97084		1 03262	1 06963	1 11215	1 16145	1 21862
2	0 92957	0 95117	0 97455		1 02795	1 05896	1 09374	1 13320	1 17830
4	0 93731	0 95675	0 97758		1 02428	1 05076	1 07988	1 11221	1 14843
6	0 94386	0 96142	0 98009		1 02133	1 04430	1 06917	1 09629	1 12609
8	0 94945	0 96537	0 98219		1 01892	1 03908	1 06066	1 08387	1 10899
4	0 95426	0 96875	0 98398		1 01690	1 03478	1 05375	1 07394	1 09551
5	0 97046	0 98000	0 98984		1 01049	1 02134	1 03256	1 04419	1 05624
6	0 97939	0 98611	0 99298		1 00718	1 01451	1 02202	1 02970	1 03756
10	0 99253	0 99500	0 99749		1 00253	1 00508	1 00764	1 01023	1 01284
25	0 99880	0 99920	0 99960		1 00040	1 00080	1 00120	1 00161	1 00201
50	0 99970	0 99980	0 99990		1 00010	1 00020	1 00030	1 00040	1 00050

THE TORONTO FUNCTION $T(\frac{1}{2}, n, r)$

r	$n = -2$	$-\frac{3}{2}$	-1	$-\frac{1}{2}$	0	$\frac{1}{2}$	1	$\frac{3}{2}$	2
0.0	Zero	$-\infty$	Zero	∞	Zero	Zero	Zero	Zero	Zero
.2	0.35845	−17.37268	0.40598	1.57656	0.54261	0.12124	0.02138	0.00321	0.00043
.4	0.50199	− 1.99641	0.57002	1.17723	0.74551	0.32333	0.11232	0.03338	0.00580
.6	0.60522	− 0.02922	0.65212	1.03988	0.87258	0.54026	0.27427	0.12035	0.04710
.8	0.68442	+ 0.50211	0.76409	0.98599	0.95080	0.73325	0.47798	0.27377	0.14030
1.0	0.74625	0.70400	0.82426	0.96741	0.99501	0.88119	0.68297	0.47250	0.29575
.2	0.79472	0.79936	0.86809	0.96445	1.01658	0.97915	0.85531	0.68100	0.49847
.4	0.83269	0.85276	0.89974	0.96778	1.02444	1.03380	0.97765	0.86176	0.70875
.6	0.86244	0.88614	0.92250	0.97292	1.02500	1.05724	1.04963	0.99264	0.89014
.8	0.88579	0.90595	0.93891	0.97794	1.02239	1.06180	1.08194	1.07014	1.01989
2.0	0.90419	0.92544	0.95086	0.98214	1.01891	1.05706	1.05878	1.10427	1.09461
.2		0.93784		0.98545		1.04911		1.11018	
.4		0.94741		0.98799		1.04106		1.10163	
.6		0.95496		0.98993		1.03415		1.08802	
.8		0.96101		0.99144		1.02861		1.07435	
3.0		0.96593		0.99262		1.02427		1.06263	
.2		0.96998		0.99357		1.02086		1.05314	
.4		0.97335		0.99434		1.01815		1.04567	
.6		0.97619		0.99498		1.01595		1.03970	
.8		0.97860		0.99552		1.01415		1.03491	
4	0.97347	0.95067	0.98816	0.99597	1.00412	1.01265		1.03199	
5	0.98283	0.98758	0.99245	0.99745	1.00258	1.00786	1.01328	1.01886	1.02461
6	0.98801	0.99136	0.99477	0.99821	1.00178	1.00538	1.00905	1.01278	1.01660
10	0.99565	0.99858	0.99812	0.99937	1.00063	1.00190	1.00317	1.00447	1.00574
25	0.99930	0.99950	0.99970	0.99990	1.00010	1.00030	1.00050	1.00070	1.00090
50	0.99983	0.99988	0.99993	0.99998	1.00003	1.00005	1.00013	1.00018	1.00023

THE TORONTO FUNCTION $T(1, n, r)$

r	$n = -2$	$-\frac{3}{2}$	-1	$-\frac{1}{2}$	0	$\frac{1}{2}$	1	-	2
0.0	One for all values of r	$-\infty$	One for all values of r	∞	One for all values of r	Zero	Zero	Zero	Zero
.1		−273.58968		5.69822		0.11246	0.00995	0.00075	0.00005
.2		−30.94617		2.93304		0.22270	0.03921	0.00588	0.00078
.3		−7.50132		2.04739		0.32363	0.08607	0.01925	0.00382
.4		−2.12570		1.63032		0.42839	0.14786	0.04378	0.01151
0.5		−0.35828		1.39928		0.52050	0.22120	0.08111	0.02650
.6		+0.34573		1.25989		0.60386	0.30232	0.13151	0.05116
.7		+0.66772		1.17157		0.67780	0.38737	0.19391	0.08719
.8		+0.82345		1.11397		0.74210	0.47271	0.26611	0.13524
.9		+0.90364		1.07578		0.79691	0.55514	0.34514	0.19481
1.0		0.94648		1.05025		0.84270	0.63212	0.42759	0.26424
.1		0.96995		1.03315		0.88021	0.70180	0.51008	0.34098
.2		0.98303		1.02171		0.91031	0.76307	0.58950	0.42190
.3		0.99040		1.01409		0.93401	0.81548	0.66334	0.50364
.4		0.99457		1.00905		0.95229	0.85914	0.72977	0.58306
1.5		0.99694		1.00575		0.96611	0.89460	0.78771	0.65745
.6		0.99828		1.00361		0.97635	0.92270	0.83678	0.72480
.7		0.99904		1.00223		0.98379	0.94442	0.87718	0.78381
.8		0.99947		1.00137		0.98909	0.96084	0.90955	0.83395
.9		0.99971		1.00082		0.99279	0.97295	0.93479	0.87529
2.0		0.99984		1.00049		0.99532	0.98168	0.95399	0.90842
.2		0.99996		1.00016		0.99814	0.99209	0.97851	0.95382
.4		0.99999		1.00005		0.99931	0.99685	0.99078	0.97870
.6		1.00000		1.00002		0.99976	0.99884	0.99636	0.99100
.8		1.00000		1.00000		0.99992	0.99961	0.99868	0.99652
3.0		1.00000		1.00000		0.99998	0.99988	0.99956	0.99877

ENTIRE FUNCTIONS BOUNDED ON A SET

By M. S. MACPHAIL

Communicated by R. L. JEFFERY, F.R.S.C.

THE following theorem was proposed by Polya [1] as a problem to be solved: *Let $f(z)$ be an entire function of minimal type of order one*, i.e., $f(z) = O(e^{\epsilon|z|})$ *for all* $\epsilon > 0$. *If the sequence* $\ldots, f(-2), f(-1), f(0), f(1), f(2), \ldots$ *is bounded, then $f(z)$ is a constant.* Proofs were given by Szegö [2], Tschakaloff [3], Paley-Wiener [4], and others. In the second and third of these proofs the points $\pm n$ figure as the zeros of $\sin \pi z$, and it is natural to ask whether these points can be replaced by the zeros z_n of a more general function $\phi(z)$. We shall state a set of conditions on $\phi(z)$ under which we can conclude that if the further requirement $f(z_n)/\phi'(z_n) = O(|z_n|^{-2})$ holds, then $f(z)$ is identically zero. Roughly speaking, any function will serve which resembles $\sin \pi z$ in that its zeros lie in a strip and it grows exponentially large away from this strip on both sides. Particular cases are Bohr's almost periodic functions and the Bessel functions of any integral order.

The conditions are as follows:

(i) $\phi(z)$ is an entire function of order one;

(ii) the zeros z_n of $\phi(z)$ all lie in a strip $|y| \leqq Y$;

(iii) there is a number q such that the zeros $z_q, z_{q+1}, \ldots$ are all simple;

(iv) there is a constant M such that $|\phi(z)| \leqq M$ in the strip $|y| \leqq Y + 1$;

(v) there are positive constants A, k such that $|\phi(x + iy)| > Ae^{k|y|}$ for $-\infty < x < \infty$, $|y| \geqq Y + 1$;

(vi) there are sequences $\{x_{r,n}\}$ $(r = 1, 2$ here and elsewhere$)$, such that $0 < x_{1,n} < x_{1,n+1} \to \infty$, $0 > x_{2,n} > x_{2,n+1} \to -\infty$, and the distance from any zero z_m to any line $x = x_{r,n}$ is bounded from zero;

(vii) the differences $|x_{r,n} - x_{r,n+1}|$ are bounded;

(viii) as a consequence of (i) and (v) the zeros z_n are infinite in number with exponent of convergence not exceeding unity [5].

We can now state the following result:

THEOREM. *Let $f(z)$ be an entire function of minimal type of order one. If $z_1, z_2, \ldots$ are the zeros of a function $\phi(z)$ satisfying the above conditions, and if the relation*

(ix) $$f(z_n)/\phi'(z_n) = O(|z_n|^{-2})$$

holds for $n \geq q$, *then* $f(z)$ *vanishes identically.*

Proof. Let $P_n(z)$ denote the principal part [5, p. 92] of $f(z)/\phi(z)$ at the point z_n; thus for $n \geq q$ we have

$$P_n(z) = \frac{f(z_n)}{\phi'(z_n)(z - z_n)}.$$

Put

(1) $$F(z) = \sum_{n=1}^{\infty} P_n(z)\phi(z).$$

By (iii), (viii), (ix), the series converges uniformly in any finite region and consequently defines an entire function [5, p. 95, sec. 2.8]. The function

$$H(z) = \frac{f(z) - F(z)}{\phi(z)}$$

is also entire, for in the neighbourhood of z_n we can write

$$H(z) = \frac{f(z)}{\phi(z)} - P_n(z) - \sum_{m \neq n} P_m(z).$$

Actually it will turn out that $H(z) = 0$, and so $f(z) = F(z)$. We have by (ii), (v), (vi),

(2) $$H(x_{r,n} + iy) = O(\exp[\epsilon\sqrt{(x_{r,n}^2 + y^2)} - k|y|]) + O(1/[1 + |y|])$$
$$= e^{\epsilon|x_{r,n}|}\, O(e^{(\epsilon - k)|y|}) + O(1/[1 + |y|]).$$

Hence,

$$\int_{-\infty}^{\infty} |H(x_{r,n} + iy)|^2\, dy = O(e^{2\epsilon|x_{r,n}|}).$$

It follows by a theorem of Hardy, Ingham and Polya [6], and (v), (vii), that

$$\int_{-\infty}^{\infty} |H(x + iy)|^2\, dy = O(e^{2\epsilon|x|}), \quad -\infty < x < \infty.$$

Next, consider the integral

$$\int_{-\infty}^{\infty} H(x + iy)\, e^{i\kappa y}\, dy.$$

On account of (v), we see as in the estimate (2) that $H(x + iy) \to 0$ as $|y| \to \infty$, uniformly in any strip; hence the above integral may be written

$$e^{-ux}\int_{-\infty}^{\infty} H(iy)e^{iuy}dy,$$

or

$$e^{-ux}h(u),\ h(u) = \int_{-\infty}^{\infty} H(iy)e^{iuy}\,dy.$$

Since for fixed x, $H(x+iy)\epsilon L^2(-\infty,\ \infty)$ as a function of y, we have by a well known theorem [7],

$$\int_{-\infty}^{\infty} |h(u)|^2 e^{-2ux}\,du = \int_{-\infty}^{\infty} |H(x+iy)|^2 dy = O(e^{2\epsilon|x|}).$$

We now wish to infer that $h(u) = 0$ almost everywhere. Suppose on the contrary this is false. Then there is a number $\kappa > 0$ such that $|h(u)| > \kappa$ on a set e of measure $\mu > 0$ (for if the sets where $|h(u)|$ is greater than 1/2, 1/3, 1/4, . . . , are all of measure zero, then $h(u) = 0$ almost everywhere). There is no loss in assuming that e lies in an interval $(a,\ b)$ with $0 < a < b$. Then for negative values of x,

$$\int_{-\infty}^{\infty} |h(u)|^2 e^{-2ux}du$$

is greater than $\mu\kappa^2e^{2a|x|}$ and so is not $O(e^{2\epsilon|x|})$ for all $\epsilon > 0$. We conclude that $h(u) = 0$ almost everywhere.

Next, by the Fourier inversion formula for functions in L^2, we find that $H(iy) = 0$ almost everywhere [7]. But since $H(z)$ is an entire function this makes $H(z)$ identically zero, and therefore $f(z) = F(z)$.

We shall now show that $F(x)\epsilon L^2(-\infty, \infty)$. Put $F(x) = F_1(x)+F_2(x)$, where $F_1(x) = \sum_1^{q-1} P_n(x)\ \phi(x)$, $F_2(x) = \sum_q^{\infty} P_n(x)\ \phi(x)$. Clearly $F_1(x)\epsilon L^2(-\infty,\ \infty)$, since $\phi(x)$ is bounded. By Cauchy's inequality,

$$|F_2(x)|^2 \leqq \sum_{n=q}^{\infty}\left|\frac{z_nf(z_n)}{\phi'(z_n)}\right|^2 \cdot \sum_{n=q}^{\infty}\left|\frac{\phi(x)}{z_n(x-z_n)}\right|^2. \tag{3}$$

We wish to integrate from $-\infty$ to ∞ under the second summation sign. Since the summands are positive and the series converges uniformly in any finite interval, this is permissible if the resulting series,

$$\sum_{n=q}^{\infty}\frac{1}{|z_n|^2}\int_{-\infty}^{\infty}|T_n(x)|^2dx,\ T_n(x) = \phi(x)/(x-z_n),$$

converges [5, p. 44].

Let C_n be the open rectangle $\Re z_n - 1 < x < \Re z_n + 1$, $|y| < Y+1$, and C_n' the boundary of C_n. Then $T_n(z)$ is analytic on $C_n + C_n''$, and by (iv), $|T_n(z)| \leqq M$ on C_n'. Hence $|T_n(z)| \leqq M$ on $C_n + C_n'$ [5, p. 165], while on the real axis outside of C_n' we have $|T_n(x)| \leqq M/|x - z_n|$. Hence

$$\int_{-\infty}^{\infty} |T_n(x)|^2\, dx \leqq 2M^2 + 2M^2 \int_1^{\infty} dx/x^2 = 4M^2.$$

It then follows by (viii), (ix) that $\int_{-\infty}^{\infty} |F(x)|^2\, dx < \infty$. Hence $f(x) = F(x) \in L_2(-\infty, \infty)$. But $f(z) = O(e^{\epsilon |z|})$, whence by a known theorem [4, p. 13, Theorem XI], we conclude that $f(z)$ is identically zero. The theorem is proved.

Almost periodic functions. We shall investigate under what conditions the analytic almost periodic functions [see 8] satisfy (i) — (viii). It is convenient to transform the theorem by rotating the axes through a right angle and to use $s = \sigma + it$ as the independent variable with s_n the zeros of $\phi(s)$; we shall, however, use the same numbering for the conditions thus modified. It will appear that in certain cases we have

(x) $$\phi'(s_n) \geqq L > 0,$$

so that (ix) takes the simpler form,

(ix′) $$f(s_n) = O(|s_n|^{-2}).$$

The result may be stated as follows.

If $\phi(s)$ satisfies the conditions described below and $f(s)$ is of minimal type of order one and satisfies (ix′), *then $f(s)$ vanishes identically.*

We now discuss the conditions to be imposed on $\phi(s)$. Assume that $\phi(s)$ is a finite sum

(4) $$\sum a_n e^{\lambda_n s}, \ \lambda_n \text{ real};$$

or more generally an entire almost periodic function with a series (4) as Dirichlet series in which the exponents λ_n are all between finite bounds. Assume further that there occur a largest and a smallest exponent

$$\Lambda = \max_n \lambda_n, \ \lambda = \min_n \lambda_n,$$

and that

(5) $$\Lambda > 0, \ \lambda < 0.$$

This last assumption is only a normalization to be obtained by multiplying by a suitable power $e^{\mu s}$. Since the bounds of $\{\lambda_n\}$ actually occur, by a known theorem [8], the zeros of $\phi(s)$ are all contained in a finite strip

(6) $$-a \leqq \sigma \leqq a.$$

Moreover, owing to (5) there exists a strip of the form (6) and positive constants A, k such that

$$|\phi(s)| \geqq Ae^{k|\sigma|} \text{ for } |\sigma| > a.$$

For all real t^* let R^* denote the closed rectangle $-a \leqq \sigma \leqq a$, $t^* - \frac{1}{2} \leqq t \leqq t^* + \frac{1}{2}$; R_0^* the open rectangle

$$-a - d < \sigma < a + d,\ t^* - \tfrac{1}{2} - d < t < t^* + \tfrac{1}{2} + d \ (d > 0);$$

and N^* the number of zeros $s_1^*, s_2^*, \ldots, s_{N^*}^*$ in R_0^*. By a lemma of Jessen [9] we have

(*a*) there is an upper bound N of the numbers N^*;

(*b*) there is a number K such that the function

$$\phi^*(s) = \phi(s) / \prod_{n=1}^{N^*} (s - s_n^*)$$

satisfies the relation $|\phi^*(s)| > K$ on R^*;

(*c*) there is a constant m such that for any point of (6) whose distance from any zero s_n of $\phi(s)$ is $\geq 1/(4n)$, the relation $|\phi(s)| > m$ holds.

It follows that there are lines $t = t_{r,n}$ satisfying conditions (v), (vi), (vii). Conditions (i), (ii), (iv) are also satisfied. Conditions (iii), (x) are satisfied provided we assume that the zeros of $\phi(s)$ are all simple and that there is a lower bound δ to the distance between any two. For then we have

$$\phi'(s_m) = \phi^*(s_m) \prod_{\substack{n=1 \\ n \neq m}}^{N^*} (s_m - s_n^*),$$

hence,
$$|\phi'(s_m)| > K\delta^N.$$

Exponential trinomials. We now consider the function $\phi(s) = ae^s + be^{ps} + ce^{-s}$, where p is irrational and $-1 < p < 1$. To show that a function of this type satisfies conditions (i)—(viii), (x), it suffices to prove that the zeros s_n are simple and that $\phi'(s_n)$ is bounded away from zero. Instead of seeking conditions under which there is a positive lower bound to the distance between any two zeros, as suggested above, we proceed as follows.

Let us first find a necessary condition for $\phi(s)$ to have a double zero. Put

$$ae^s + be^{ps} + ce^{-s} = 0, \tag{7}$$

$$ae^s + bpe^{ps} - ce^{-s} = 0. \tag{8}$$

Adding (7) and (8),

$$2ae^{s}+b(\rho+1)e^{\rho s}=0,$$

$$e^{(\rho-1)s}=-\frac{2a}{b(\rho+1)},$$

$$e^{(\rho-1)\sigma}=\frac{2|a|}{|b|\,|\rho+1|},$$

$$e^{\sigma}=\left\{\frac{2|a|}{|b|\,|\rho+1|}\right\}^{1/(\rho-1)}.$$

Subtracting (8) from (7),

$$b(1-\rho)e^{\rho s}+2ce^{-s}=0,$$

$$e^{(\rho+1)s}=\frac{2c}{b(\rho-1)},$$

$$e^{(\rho+1)\sigma}=\frac{2|c|}{|b|\,|\rho-1|},$$

$$e^{\sigma}=\left\{\frac{2|c|}{|b|\,|\rho-1|}\right\}^{1/(\rho+1)},$$

whence,

$$\frac{|b|^{2}|c|^{\rho-1}}{4|a|^{\rho+1}}=\frac{|\rho-1|^{\rho-1}}{|\rho+1|^{\rho+1}}. \tag{9}$$

Suppose now a, b, c do not satisfy (9). Then $\phi(s)$ has no double zero. It is easily checked (without necessarily referring to [8]) that the zeros $s_n=\sigma_n+it_n$ of $\phi(s)$ all lie in a strip $-\alpha\leqq\sigma\leqq\alpha$. Suppose if possible that $\phi'(s_n^{(1)})\rightarrow 0$ on a certain subsequence $s_n^{(1)}$ of s_n. Let

$$\phi_n^{(1)}(s)=\phi(s+it_n^{(1)})=a_n^{(1)}e^{s}+b_n^{(1)}e^{\rho s}+c_n^{(1)}e^{-s},$$

where

$$a_n^{(1)}=a_ne^{it_n^{(1)}},\quad b_n^{(1)}=b_ne^{i\rho t_n^{(1)}},\quad c_n^{(1)}=c_ne^{-it_n^{(1)}}.$$

Then each function $\phi_n^{(1)}(s)$ has a zero on the real axis in the interval $a\leqq\sigma\leqq\beta$. Choose a subsequence $t_n^{(2)}$ of $t_n^{(1)}$ such that $a_n^{(2)}$, $b_n^{(2)}$, $c_n^{(2)}$ converge to A, B, C respectively. Then the corresponding subsequence $\phi_n^{(2)}(s)$ converges uniformly to $\Phi(s)=Ae^{s}+Be^{\rho s}+Ce^{-s}$. Choose a subsequence $t_n^{(3)}$ of $t_n^{(2)}$ such that the zeros $\sigma_n^{(3)}$ of $\phi_n^{(3)}(s)$ tend to a limit point σ_0. Then by uniform convergence,

$$\Phi_n^{(3)}(\sigma_n^{(3)})\rightarrow\Phi(\sigma_0),\quad \Phi_n^{(3)\prime}(\sigma_n^{(3)})\rightarrow\Phi'(\sigma_0).$$

It follows that $\Phi(\sigma_0) = \Phi'(\sigma_0) = 0$. But since $|A| = |a|$, $|B| = |b|$, $|C| = |c|$, A, B, C do not satisfy (9), and so $\Phi(s)$ cannot have a double zero.

This contradiction shows that $\phi'(s_n)$ is bounded away from zero. Hence the result:

Assume that $\phi(s)$ *is of the form* $ae^s + be^{\rho s} + ce^{-s}$, *with* ρ *irrational*, $-1 < \rho < 1$, *and*

$$\frac{|b|^2|c|^{\rho-1}}{4|a|^{\rho+1}} \neq \frac{|\rho - 1|^{\rho-1}}{|\rho + 1|^{\rho+1}}.$$

If $f(s)$ *is of minimal type of order one, and* $f(s_n) = O(|s_n|^{-2})$, *where* $\{s_n\}$ *are the zeros of* $\phi(s)$, *then* $f(s)$ *vanishes identically.*

Bessel functions. In this section we shall take $\phi(z)$ to be a Bessel function[1] $J_m(z)$ of integral order m. We shall show that the theorem may be modified to yield the following result:

If a function $f(z)$ *of minimal type of order one is bounded at the zeros of* $J_m(z)$, *it is a constant.*

We may assume $f(z)$ to be an even function. For suppose we have proved the theorem on this assumption; if now $f(z)$ is odd it will follow that the even function $f(z)/z$ is a constant which can only be zero. But any function $f(z)$ can be written as the sum of the functions $[f(z) + f(-z)]/2$, $[f(z) - f(-z)]/2$, which are even and odd respectively. Since the zeros of a Bessel function are symmetrical with respect to the origin, the condition $f(z_n) = O(1)$ holds also for the even and odd parts of $f(z)$.

Suppose then $f(z)$ is even and write

$$g(z) = \frac{f(z) - f(0)}{z^2}.$$

Put $z_0 = 0$ and let $z_{\pm n}$ denote the remaining zeros of $J_m(z)$. Instead of (1) and (3) write

$$(1') \qquad G(z) = P_0(z)J_m(z) + \sum_{n=-\infty}^{\infty}{}' \frac{g(z_n)J_m(z)}{J'_m(z_n)(z - z_n)}$$

and

$$(3') \qquad G_2(x)|^2 \leqq \sum_{n=-\infty}^{\infty}{}' \left|\frac{z_n^{\frac{1}{2}}g(z_n)}{J'_m(z_n)}\right|^2 \sum_{n=-\infty}^{\infty}{}' \left|\frac{J_m(z)}{z_n^{\frac{1}{2}}(x - z_n)}\right|$$

where Σ' indicates that the term given by $n = 0$ is omitted. To establish the convergence of the series we shall show that

---

[1] The results here used may be found in the treatises of G. N. Watson or Gray and Mathews.

$1/J'_m(z_n) = O(|z_n|^{\frac{1}{2}})$. We have

$$zJ''_m(z) = nJ_m(z) - zJ_{m+1}(z),$$

whence

$$J''_m(z_n) = -J_{m+1}(z_n). \tag{10}$$

Also for $p = m, m+1$ we have

$$J_p(z) = \sqrt{[2/(\pi z)]}\,[\cos(z - p\pi/2 - \pi/4) + \theta_p(z)], \tag{11}$$

where $|\theta_p(z)| \leq 1/10$ provided $|z| \geq Z$, say. Since this does not hold on the negative real axis, we may at first take $\Re z \geq 0$, and later make use of the equation $J_m(-z) = (-1)^m J_m(z)$. The equation $1/J'_m(z_n) = O(|z_n|^{\frac{1}{2}})$ is easily deduced from (10), (11), noting that there are only a finite number of zeros with moduli less than Z. The existence of the lines $x = x_{r,n}$ also follows easily. We have

$$|J_m(x_{r,n} + iy)| > Ae^{k|y|}\,|y|^{-\frac{1}{2}}$$

instead of

$$|J_m(x_{r,n} + iy)| > Ae^{k|y|},$$

but it is seen on examining (2) that this does not impair the proof. The desired result for Bessel functions now follows.

Princeton University.

REFERENCES

(1) Jahresber. deut. Math. Verein., **40**, 2nd part: 80, problem 105. 1931.
(2) *Ibid.*, **43**, 2nd part: 10. 1934.
(3) *Ibid.*: 11.
(4) Paley, R. E. A. C., and Wiener, N. Fourier Transforms, 81. New York, 1934.
(5) Titchmarsh, E. C. The Theory of Functions, 250. Oxford, 1932.
(6) Hardy, G. H., Ingham, A. E., and Polya, G. Theorems concerning mean values of analytic functions. Proc. Royal Soc. London, A, **113**: 542-567, 554. 1927.
(7) Zygmund, A. Trigonometrical Series, 315, Theorem (ii). Warsaw, 1935. The result can also be deduced from Titchmarsh, *op. cit.*: 436-439, and 443, example 21.
(8) Besicovitch, A. S. Almost periodic functions. 162. Cambridge, 1932.
(9) Jessen, B. Uber die Nullstellen einer analytischen fastperiodischen Funktion. Math. Ann., **108**: 485-516, 499. 1933.

THE COLOURS OF SNOWFIELDS IN SUNLIGHT[1]

By W. E. KNOWLES MIDDLETON, F.R.S.C.

IN 1935 the writer worked out a theory (1) of the apparent colours of distant objects, produced by the action of the atmosphere on light. In the calculations with which the theory was demonstrated, it was assumed that the sky was completely and uniformly clouded (except in the case of black objects), the reason for this assumption being that it made possible relatively exact numerical results. It was shown that under such conditions a white object appears white at any distance, though it is actually a blue tint too pale to be distinguished from white.

In a private communication, Dr. H. Tschierske reminded the writer of the observations of A. Heim, so beautifully set forth in his book, *Luft-Farben* (2); in particular of Heim's remarks on the yellow and orange tints of distant snowfields lit by the sun, even, it would appear, at mid-day. This made it seem desirable to make some calculations of the apparent colours of distant white objects in bright sunlight, using the previous theory (1) and making reasonable assumptions about the illumination. The reader must be referred to the above-mentioned paper for a description of the method of colour specification, as it would take too much space to repeat it here.

Let us suppose a mountain slope facing the sun, and covered with freshly fallen snow. Dorno (3) measured the brightness of one such surface and found it to be 5 times that of the sky near it. This was, however, 21° above the true horizon; let us therefore, to be on the safe side, assume an upper limit of 4 times the horizon brightness.

Let the ratio (intrinsic brightness[2] of object/brightness of horizon sky) be denoted by *k*. Since such a surface will be illuminated by the sun and also by most of the sky, we shall not be far wrong if we assume the light incident on it (and reflected from it, since it is non-selective) to have an energy distribution similar to C.I.E. Source "C" (4). The air between it and the observer will be illuminated by the same light. We have therefore to add the light scattered towards the observer by the air,

$$H_s = H_h(1 - e^{-\sigma l}) \qquad \text{........(1)},$$

to the light

$$H'_w = kH_h e^{-\sigma l} \qquad \text{........(2)},$$

---

[1] Published by permission of the Controller of Meteorological Services, Department of Transport, Canada.

[2] I.e., the brightness which would be observed close to the object.

reaching the eye from the object. Here H_h is the brightness of the horizon sky, σ the extinction coefficient of the atmosphere in the horizontal per km, and l the distance of the object in kilometres. Considering for the moment only a narrow band of the spectrum about the wavelength λ, we may write for the energy of this wavelength reaching the eye

$$H_{w,\lambda} = H_{h,\lambda}(1 - e^{-\sigma_\lambda l}) + kH_{h,\lambda}e^{-\sigma_\lambda l} =$$
$$H_{h,\lambda}[1 + (k - 1)e^{-\sigma_\lambda l}] \qquad \ldots\ldots..(3),$$

where σ_λ is the extinction for light of wavelength λ.

In pure air or in haze, σ will vary with λ. In a sample of very light haze measured (5) by the writer, it was found that $\sigma_\lambda = 0.0080\lambda^{-2.09}$ km^{-1}. A black object could be seen against the sky for about 150 km in such air, which we shall use in the following calculations.

TABLE I

APPARENT COLOURS OF A SNOWFIELD IN SUNLIGHT

Assuming illumination equivalent to C.I.E. Source "C", and a haze in which $\sigma = 0.0080\lambda^{-2.09}$

Assumed intrinsic brightness ($H_h = 1$)	Distance l km	Apparent brightness H_w ($H_h = 1$)	Trichromatic coefficients x	y	z	Hue wave-length λ_s microns	Saturation per cent S	p_e
4.00	20	2.738	.3238	.3287	.3475	0.579	10.5	7.0
	30	2.328	.3365	.3403	.3232	.579	19.6	13.5
	40	2.014	.3409	.3430	.3161	.580	22.3	15.7
	50	1.777	.3420	.3426	.3154	.581	22.7	16.2
	100	1.208	.3316	.3300	.3384	.584	13.2	9.5
	150	1.057	.3191	.3204	.3005	.589	4.7	3.5
3.00	20	2.159	.3216	.3268	.3516	0.579	8.9	5.9
	30	1.885	.3315	.3358	.3327	.579	10.2	11.0
	40	1.677	.3343	.3374	.3283	.580	17.9	12.4
	50	1.518	.3345	.3364	.3291	.580	17.1	12.0
	100	1.139	.3250	.3256	.3494	.584	9.1	6.6
	150	1.038	.3162	.3191	.3647	.588	3.1	2.3
2.00	50	1.259	.3245	.3281	.3474	0.580	10.4	7.0

Using equations (3) of the previous paper, and employing H_h as the unit of luminosity, we have for the tristimulus values of the apparent colour of the snowfield

$$\left.\begin{aligned} x' &= \int_0^\infty E_\lambda \bar{x}_\lambda \left[1 + (k-1)e^{-\sigma_\lambda l}\right] d\lambda \\ y' &= \int_0^\infty E_\lambda \bar{y}_\lambda \left[1 + (k-1)e^{-\sigma_\lambda l}\right] d\lambda \\ z' &= \int_0^\infty E_\lambda \bar{z}_\lambda \left[1 + (k-1)e^{-\sigma_\lambda l}\right] d\lambda \end{aligned}\right\} \quad \ldots\ldots\ldots(4).$$

The trichromatic coefficients then are

$$x = \frac{x'}{x' + y' + z'},\; y = \frac{y'}{x' + y' + z'},\; z = \frac{z'}{x' + y' + z'} \quad \ldots\ldots\ldots(5).$$

These calculations have been performed for snowfields (a) 4 times (b) 3 times as bright as the horizon sky, and at distances 20, 30, 40, 50, 100, and 150 km. They were carried out by the method of "selected ordinates" described by Hardy (4), which is more convenient than the method used by the writer in his previous work. The results are given in Table I.

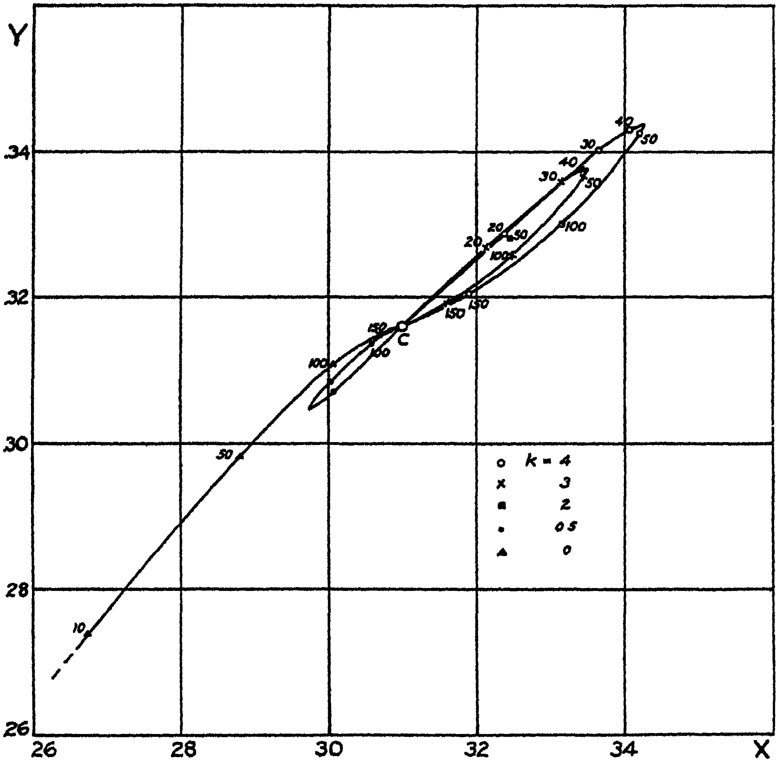


FIGURE 1

In addition to the trichromatic coefficients, the hue-wavelength and saturation have been found for each example. The saturation is given for convenience in two ways - S is the saturation as defined in the previous paper (1) (called *colorimetric purity* by Hardy (4)); p_e is the *excitation purity* (Hardy).

These results are shown graphically in Fig. 1. The zx and zy axes, now universally employed, have been used instead of the xz and xy axes used in the previous paper.

It will be obvious that when the distance is zero or infinite the apparent colour of the object will be that of the illumination. The calculations show, however, that the object appears yellower with increasing distance up to some 40 or 50 km after which the yellow becomes paler. The hue-wavelength remains near 0.585μ at all distances.

The more brightly the snowfield is illuminated, the yellower it will appear at a given distance. The loop formed by the curve shrinks to a point (at *C*) for $k = 1$; this means that a snowfield actually of the same brightness as the horizon will not change its apparent colour as we recede from it. It has not been considered necessary to calculate curves for any more values of k, but the point for $k = 2.00$, $l = 50$ km has been found, and appears on the diagram and in the table.

It is interesting to compare these curves with those for black and white objects under a cloudy sky, if we remember that for the white object under a cloudy sky, $k = 0.5$, and for the black object $k = 0$. The curve for the white object is a loop lying on the opposite side of "C" to the present curves, and is also shown, as is the curve for the black object. Thus the entire family of curves, as k varies from 0 to 4, forms a shape much like a two-bladed propellor, with the hub at the position of source "C".

It is of course difficult to predict the actual appearance of these colours from the co-ordinates, since the system has been established under laboratory conditions, and not under those found in the open air (6). However, there is little doubt that all of them which correspond to distances between 30 and 100 km will be seen as yellow, at least by many observers, though the results of Martin, Warburton, and Morgan (7) suggest that people differ widely in their ability to distinguish pale colours from white, especially in the yellow.

The atmosphere may, on rare occasions, be more selective than the sample on which these calculations were based. In this event the colours will be more saturated. It appears that the necessary conditions for the yellow appearance of distant snowfields are, first, that they should be brightly illuminated by the sun; and second, that the atmosphere should be rather clear or at any rate highly selective.

There is one further result that should be mentioned. In air with these properties, the black, grey, and white objects referred to in the previous paper all reached the threshold of contrast (H differing from H_h by 2 per cent) at distances a little less than 150 km. The objects now being considered have not reached this threshold at 150 km; indeed, the object for which $K = 4$ will not do so until a distance of about 170 km is attained. This result is in accord with experience, since a snowy peak or a brightly illuminated cloud is often the only object visible at an extreme distance, rising ghost-like out of nothing. But this is not conclusive evidence of the correctness of the theory, since the generally greater elevation of the peak or the cloud will also operate to increase its visual range.

REFERENCES

(1) MIDDLETON, W. E. K. On the colours of distant objects, and the visual range of coloured objects. Trans. Royal Soc. Canada, sec. III, **29**: 127-154. 1935.

(2) HEIM, A. Luft-Farben. Zurich 1912, Hofer and Co. A.G.

(3) DORNO, C. Himmelshelligkeit, Himmelspolarisation und Sonnenintensitat in Davos 1911 bis 1918. Veröff. d. Preuss. Met. Inst., no. 303, 1919, Table 50b.

(4) HARDY, A. C. Handbook of Colorimetry. Cambridge, Mass., The Technology Press, p. 49. 1936.

(5) MIDDLETON, W. E. K. Experiments with a telephotometer—the dependence of extinction coefficient on wavelength. Gerlands Beitr. zur Geophys., **44**: 358-375. 1935.

(6) JUDD, D. B. Surface color. J.O.S.A., **25**: 44. 1935.

(7) MARTIN, L. C., F. L. WARBURTON, and W. J. MORGAN. Determination of the sensitiveness of the eye to differences in the saturation of colours. Medical Research Council, Reports of the Committee upon the Physiology of Vision, XIII. Special Report Series, no. 188, London. 1933.

ON A STATISTICAL MODEL OF ISOTROPIC TURBULENCE

By J. L. SYNGE, F.R.S.C., and C. C. LIN

ABSTRACT

A statistical theory of homogeneous isotropic turbulence is developed by regarding the turbulent fluctuations as due to Hill's vortices moving about chaotically. The calculated velocity correlation tensor of the second order satisfies the general relations of von Kármán, and the shape of the correlation function is in general agreement with experimental results.

1. *Introduction*

The kinetic theory of gases is perhaps the best example of a physical theory in which a definite microscopic model is made to yield macroscopic results. The crude microscopic model, consisting of elastic spheres, has proved remarkably successful. To construct a microscopic picture of turbulence, we do not, of course, have to go down to the molecular scale, but we do have to consider a state of motion so chaotic that the classical equations of hydrodynamics must be averaged. This process of averaging, as usually performed, is by no means clear, and in this and other respects existing theories of turbulence compare unfavourably with the kinetic theory of gases in logical precision.

Our purpose in the present paper is to offer a microscopic theory of turbulence with some claim to logical completeness. It is based on a model of the turbulent regime in which the individual eddies are treated as spherical vortices. If this model appears over-simplified, we may defend it on the ground that it is successful in giving results, at least as far as the deductions of the present paper are concerned. We find correlation functions which satisfy the conditions of von Kármán, and are in good agreement with observation.

As a background, we should mention the outstanding contributions to the theory of isotropic turbulence. The first theoretical study was by G. I. Taylor in 1921 (1) and a clear formulation of the concept was given by him in 1935 (2). A systematic general theory was given by von Kármán and Howarth in 1937-8 (3); this was simplified by Robertson (4), who used the theory of invariants. Recently a theory of isotropic turbulence in two dimensions has been formulated by Burgers(5).

Although the properties of a model consisting of concentrated vortices does not seem to have been systematically investigated before, we may remark that G. I. Taylor (2) emphasized the importance of small eddies in a turbulent field; he has regarded the length λ derived from correlation measurements as a measure of the diameters of the smallest eddies. He and Green (6) have further given a model system illustrating the mechanism of the production of small eddies from large ones.

Before proceeding to the mathematical details of our theory, let us form a general picture of the formation of turbulence behind a grid in a wind tunnel. As the air blows through the grid, vorticity is imparted to the portions of the fluid which pass close to the bars of the grid. On account of the instability of continuous vortex distributions, the portions of the fluid endowed with vorticity tend to be concentrated into eddies. Except for slow dissipation due to viscosity, these eddies carry their vorticity downstream, and at some distance behind the grid we have the physical state with which this paper deals in an idealized form. We shall regard the regime as one of *isotropic and homogeneous turbulence*. Isotropy means that all orientations of the eddies are equally probable and homogeneity means that the statistical volume density of eddies is constant throughout the region under consideration.

In practice, various influences keep the regime from being strictly isotropic and homogeneous. Thus, the grid produces vortex rings with a definite orientation, and this will not be entirely lost. Further, the walls of the tunnel produce some lack of homogeneity, and perhaps of isotropy. Finally, as we pass down-stream, the effects of viscosity produce changes. Consequently, our assumption of isotropy and homogeneity is in the nature of an idealization, but it is the idealization usually made, and we are not called on to defend it as in any way a novel assumption.

In creating a microscopic model of isotropic and homogeneous turbulence, we have had two requirements in mind:

(i) the model must resemble fairly closely the actual physical state;

(ii) it must be sufficiently simple for mathematical use.

In our first attempt, we represented the eddies by infinitesimal vortex rings, the velocity-field due to each being that of a doublet. This model, however, has undesirable features and does not lead to correlation functions in agreement with observation. It is, in fact, necessary to use eddies of finite size, and we finally decided to use Hill's spherical vortices (7), which prove successful. As an alternative, we have considered finite vortex rings, but the mathematical calculations are unmanageable.

The Hill vortex is not a perfect model. In an inviscid fluid, such a vortex moves steadily without change of type, but across its surface there is a discontinuity in tangential rate of shear. Hence, if the fluid is viscous, a discontinuity in shearing stress is called into play, which must quickly alter the motion. We would have a better model if we could find a fairly simple axially symmetric steady motion of a perfect fluid, without this discontinuity. But such a motion does not seem to be known.

2. *Statistical representation of a turbulent field*

Let us now apply our model to the calculation of correlation functions. For this purpose, we must know two things: first, what the properties of the individual vortices are, and secondly, how to calculate average values.

As mentioned above, we accept Hill's spherical vortex as the model of an eddy. Each vortex is characterized by two quantities, its speed U and its radius a. We shall not suppose that these vortices are all the same. On the contrary, there will be a statistical distribution in respect of type, to be discussed later. The quantities (U, a) define a type.

The calculation of average values will now be explained. We fix our attention on a definite portion of space R behind the grid, fixed relative to the tunnel. (If we observed a portion of the fluid carried down-stream with the mean velocity, we would not observe a steady regime, because of the effect of viscosity.) We imagine an observer, who can take instantaneous microscopic pictures of R; such a picture will be called an *observation*. Each observation shows a discrete distribution of vortices in R. When a very great number of such observations are superimposed, we get an ensemble, in which the various quantities have (in the limit) continuous statistical frequency distributions.

We have now to make two simplifying assumptions. First, since vortices move in and out of R, their number will not remain the same throughout a sequence of observations. But if the number of vortices is large, this random fluctuation is of no statistical importance, and we shall for simplicity assume the number of vortices in R to remain a constant number N. Secondly, we shall feel free to regard R as of infinite extent, when we find it convenient to do so. (Physically, this means that R is large in comparison with the distance between the points between which we calculate correlations.)

Let us now proceed to the mathematical formulation of the problem. In any particular observation, we may label the N vortices present from 1 to N. This can evidently be done in $N!$ ways. To describe each vortex, we require seven quantities: its size a, the Cartesian coordinates of its centre x_i $(i = 1, 2, 3)$, and the components of velocity U_i $(i = 1, 2, 3)$. To describe the observation completely, we can therefore use a point in the $7N$-dimensional phase-space with coordinates $(a_1, a_2, \ldots, a_m, \ldots, a_N)$, where a_m denotes the set of 7 variables $((x_1)_m, (x_2)_m, (x_3)_m, (U_1)_m, (U_2)_m, (U_3)_m, a_m)$ of the vortex carrying the number m. Since there are $N!$ ways of numbering the vortices, there are $N!$ different points in the phase-space which correspond to a given observation.

On superimposing a very great number of observations, we get a (statistically) continuous frequency distribution in the phase-space. Let

$$\phi(a_1, a_2, \ldots, a_N)$$

be the probability density function; this means that the frequency of occurrence of the microscopic state in which the m-th vortex is in the range of states $(a_m, a_m + da_m)$ is given by

$$\phi(a_1, a_2, \ldots a_N)da_1da_2 \ldots da_N, \tag{2 1}$$

where

$$\int \ldots \int \phi(a_1, a_2, \ldots, a_N)da_1da_2 \ldots da_N = 1, \tag{2.1a}$$

the integral being taken for the whole of the phase-space.

Evidently, the order of numbering is immaterial and therefore ϕ is symmetric in all the sets of variables $a_1, a_2, \ldots, a_N$. The ensemble average of any quantity $Q(a_1, \ldots, a_N)$ is then

$$\overline{Q} = \int \ldots \int Q\phi \, da_1 \ldots da_N. \tag{2.2}$$

In accordance with the usual hypothesis of statistical mechanics, we identify this ensemble average with the average obtained from a great number of actual observations.

In the field of turbulence, let us now fix our attention on two points A_0 and $\tilde{A}_0$, for which the correlation functions are to be calculated (Fig. 1). Let $p(a_m)$ and $\tilde{q}(a_m)$ be the contributions of the vortex m (at the point A) to two quantities P and $\tilde{Q}$ at the points A_0 and $\tilde{A}_0$ respectively. These quantities are, for the purpose of the present problem, components of velocity, but the argument merely demands that they are to be calculated by adding the contributions from the several vortices; thus

$$P = \sum_{m=1}^{N} p_m, \quad \tilde{Q} = \sum_{m=1}^{N} \tilde{q}_m, \tag{2.3}$$

where $p_m = p(a_m)$, $\tilde{q}_m = \tilde{q}(a_m)$.

Then, by (2.2), we have

$$\overline{P\tilde{Q}} = \int \cdots \int \sum_{1}^{N} p(a_m) \sum_{1}^{N} \tilde{q}(a_n)\phi\, da_1 \ldots da_N \tag{2.4}$$

$$= \sum_{m,n} \int \cdots \int p(a_m)\, \tilde{q}(a_n)\phi\, da_1 \ldots da_N$$

$$= \sum_{m,n} \overline{p_m \tilde{q}_n}\,.$$

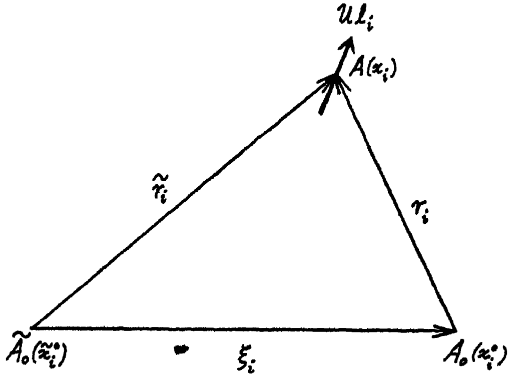


FIGURE 1.—Schematical diagram for the calculation of velocity correlations due to a Hill's vortex.

U = speed of the vortex,
A = centre of the vortex,
$\tilde{A}_0, A_0$ = points where the fluctuations are observed.

The calculation of average values is impossible unless we know the function ϕ. We can, however, make a great simplification by analogy with the kinetic theory of gases. In an ideal gas, the molecules behave independently. We shall assume this to be so with the vortices. Furthermore, in an isotropic homogeneous field of turbulence, there are no preferential directions and positions. We assume the vortices to be distributed and oriented at random without any correlation as regards type, position and orientation.

This means that while a certain vortex V_1 is occupying a position A with a certain orientation in the field, another vortex V_2 can (with equal probability) occupy any position B with arbitrary orientation. In fact, we are not excluding the mathematical possibility of having the vortex V_2 at the position A. This is not true physically, of course, for the vortices have finite sizes and cannot overlap without destroying their structure. The theory is therefore to be regarded as a first approximation, just as the ideal gas theory is a first approximation to Van der Waals' theory.

Mathematically, the assumption of the independence of the vortices from one another, combined with the fact that the order of numbering is immaterial, gives

$$(2.5) \qquad \phi(a_1, a_2, \ldots, a_N) = \psi(a_1)\psi(a_2)\ldots\psi(a_N),$$

with

$$(2.5a) \qquad \int\psi(a_m)da_m = 1, \ (m = 1, 2, \ldots, N),$$

the integral being taken over the 7-dimensional space of x_i, U_i, a. Thus,

$$(2.5b) \qquad \begin{cases} \overline{p_m \tilde{q}_m} = \int p(a_m)\tilde{q}(a_m)\psi(a_m)da_m \\ \overline{p_m \tilde{q}_n} = \int p(a_m)\psi(a_m)da_m \int q(a_n)\psi(a_n)da_n, \ (m \neq n) \end{cases}$$

and so

$$(2.6) \qquad \overline{p_m \tilde{q}_n} = \bar{p}_m \bar{\tilde{q}}_n, \text{ if } m \neq n.$$

Hence

$$(2.7) \qquad P\tilde{Q} = P\,\tilde{Q} + \sum_{m=1}^{N} (p_m \tilde{q}_m - p_m \tilde{q}_m).$$

We shall suppose that P, $\tilde{Q}$ are two "fluctuating" quantities, so that

$$(2.8) \qquad P = 0, \quad \tilde{Q} = 0.$$

From this assumption it follows that

$$(2.8a) \qquad p_m = 0, \ \tilde{q}_m = 0, \quad (m = 1, 2, \ldots, N),$$

since, by (2.3),

$$P = Np_m, \quad \tilde{Q} = N\tilde{q}_m \, (m = 1, 2, \ldots, N).$$

Then (2.7) reduces to

$$(2.9) \qquad \overline{P\tilde{Q}} = \sum_{m=1}^{N} \overline{p_m \tilde{q}_m} = Np\tilde{q},$$

where p, $\tilde{q}$ are the contributions due to a single typical vortex. It will no longer be necessary to use a suffix to indicate a particular vortex. The calculation of the average value $\overline{P\tilde{Q}}$ can therefore be made by considering a typical vortex, and calculating the average

$$\overline{p\bar{q}} = \int p(a)\bar{q}(a)\psi(a)da. \tag{2.10}$$

On account of the independence of position, orientation and type, we may write

$$\psi(a)da = \mu(l)d\Omega \,.\, \sigma(x)d\tau \,.\, \Phi(U,a)dUda. \tag{2.11}$$

Here $d\Omega$ is an element of solid angle, l stands for the direction cosines l_i of the axis of the vortex, and μ is frequency of distribution in direction; $d\tau$ is an element of spatial volume of the region R under consideration, x stands for the coordinates x_i of the centre of the vortex, and σ is frequency of distribution in position; U is the speed and a the radius of the vortex, and $\Phi(U, a)$ is frequency of distribution in type. We have

$$\int \mu(l)d\Omega = 1, \quad \int \sigma(x)d\tau = 1, \quad \int \Phi(U, a)dUda = 1. \tag{2.11a}$$

The first integral is taken over a unit sphere, the second over the interior of the region R, and the third for all positive values of U and a. If V be the volume of the space-region R under observation, the conditions of isotropy in direction, and homogeneity in position, give

$$\mu = \frac{1}{4\pi}, \quad \sigma = \frac{1}{V}. \tag{2.11b}$$

Introducing (2.11) into (2.10), we have

$$\overline{p\bar{q}} = \frac{1}{V}\int \Phi(U, a)dUda \int_V \overline{p\bar{q}}^{\,l}\, d\tau, \tag{2.12}$$

where $\overline{p\bar{q}}^{\,l}$ is the average taken for orientation only, and of course depends on x_i and U and a. Then $\overline{P\bar{Q}}$ is given (cf. (2.9)) on multiplying (2.12) by N.

There is no objection to extending the volume V to infinity while keeping the density $\rho = N/V$ of the vortices constant, provided that the volume integral is convergent. Thus, we get

$$\overline{P\bar{Q}} = \rho \int \Phi(U, a)dUda \int \overline{p\bar{q}}^{\,l}\, d\tau, \tag{2.13}$$

the volume integral being taken through infinite space.

The treatment of correlations of more than two quantities is similar. Thus,

$$\overline{PQ\bar{R}} = \rho \int \Phi(U, a)dUda \int \overline{pq\bar{r}}^{\,l}\, d\tau, \tag{2.14}$$

$$PQ\tilde{R}\hat{S} = \rho\int\Phi(U,a)\,dU\,da\int p\,q\,\tilde{r}\,\hat{s}'\,d\tau, \text{ etc.} \qquad (2.15)$$

It may be noted in passing that if all the vortices are of the same type, the average values are proportional to the volume integrals themselves.

The formula (2.13) forms the basis of our calculations.

3. *Hill's spherical vortex*

Hill's spherical vortex may be described as a vortex sphere advancing with constant velocity through a fluid which is at rest at infinity. The flow in the whole field is symmetrical about the line of motion. The vorticity inside the sphere is proportional to the distance from this axis of symmetry, while the motion outside is the same as the irrotational motion outside a moving solid sphere.

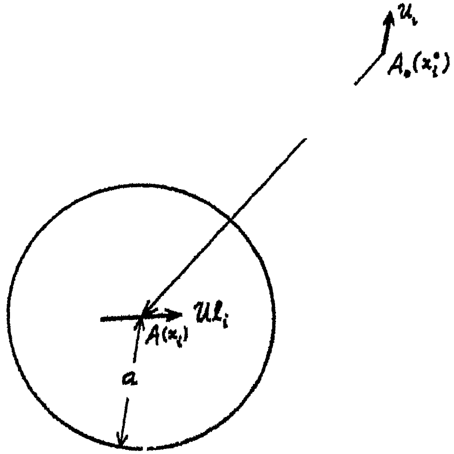


FIGURE 2.—Hill's spherical vortex.

In Fig. 2, $A(x_i)$ $(i = 1, 2, 3)$ is the instantaneous position of the centre of a Hill's vortex of radius a, and l_i $(i = 1, 2, 3)$ are the direction cosines of the direction of advance referred to any coordinates axes (not shown). Let U be the speed of advance. Then the fluid velocity at a point $A_0(x_i^0)$ is given by

$$u_i = \alpha_{ij}l_j, \quad \alpha_{ij} = \frac{3}{2}\,\frac{Ua^3}{r^5}\,r_ir_j - \frac{1}{2}\,\frac{Ua^3}{r^3}\,\delta_{ij}, \quad (r \geq a), \qquad (3.1a)$$

$$u_i = \beta_{ij}l_j, \quad \beta_{ij} = \frac{3}{2}\,\frac{U}{a^2}\,r_ir_j + U\left(\frac{5}{2} - \frac{3r^2}{a^2}\right)\delta_{ij}, \quad (r \leq a), \qquad (3.1b)$$

where $r_i = x_i - x_i^0$, the coordinates of A relative to A_0. These are well-known formulae, translated into tensor notation. As usual, r denotes the length of the vector r_i, and δ_{ij} is the Kronecker delta, while summation over the range (1, 2, 3) for a repeated index is understood. The pressure is given by

$$(3.2a)\qquad p = \Pi + \frac{\rho U^2}{8}\left\{\left[5 - 4\left(\frac{a}{r}\right)^3 - \left(\frac{a}{r}\right)^6\right] + 3\left[4\left(\frac{a}{r}\right)^3 - \left(\frac{a}{r}\right)^6\right]\frac{(r_i l_i)^2}{r^2} + \frac{9}{4}\right\},\quad (r \geq a),$$

$$(3.2b)\qquad p = \Pi + \frac{9\rho U^2}{8a^4}\left\{\left[r^2 - (r_i l_i)^2 - \frac{1}{2}a^2\right]^2 - \left[(r_i l_i)^2 - a^2\right]^2 + a^4\right\},\quad (r \leq a)$$

where Π is a constant (being in fact the minimum pressure in the fluid).

It is easy to verify that these fields of velocity and pressure satisfy the equations of steady motion of a perfect fluid, when a velocity $-Ul_i$ is superimposed and x_i regarded as fixed. Further, the velocity and pressure are continuous across $r = a$. The continuity of velocity follows from the fact that

$$(3.3)\qquad \alpha_{ij} = \beta_{ij} \text{ at } r = a,$$

while the continuity of pressure is evident from (3.2a) and (3.2b).

We note that the tensors α_{ij} and β_{ij} are obviously symmetrical, and we can easily verify that their divergence with respect to either index vanishes:

$$(3.4)\qquad \frac{\partial \alpha_{ij}}{\partial x_i} = 0,\quad \frac{\partial \alpha_{ij}}{\partial x_j} = 0,\quad \frac{\partial \beta_{ij}}{\partial x_i} = 0,\quad \frac{\partial \beta_{ij}}{\partial x_j} = 0.$$

These also follow directly from the equation of continuity $\dfrac{\partial u_i}{\partial x_i^0} = 0$, since α_{ij}, β_{ij} are independent of l_k.

It is evident that we can write

$$(3.5)\qquad \alpha_{ij} = \frac{\partial \varphi_i}{\partial x_j},\quad \varphi_i = \frac{1}{2}Ua^3\left(-\frac{r_i}{r^3}\right).$$

The same reduction is not possible for β_{ij}, because the motion inside is rotational. However, we may express β_{ij} in the form

$$(3.6)\qquad \left\{\begin{aligned} {}_{ij} &= \frac{\partial \psi_i}{\partial x_j} + q(r)\delta_{ij},\quad \psi_i = \frac{3}{4}\frac{U}{a^2}r^2 r_i - \frac{5}{4}Ur_i,\\ q &= -\frac{\partial \psi_i}{\partial x_i} = -\frac{15U}{4a^2}(r^2 - a^2).\end{aligned}\right.$$

It is further possible to write

$$\begin{cases} \varphi_i = \dfrac{\partial u}{\partial x_i}, \quad u = \dfrac{1}{2}\, U \dfrac{a^3}{r}, & (3.7) \\[2ex] \psi_i = \dfrac{\partial v}{\partial x_i}, \quad v = \dfrac{3}{16}\dfrac{U}{a^2} r^4 - \dfrac{5}{8}\, U r^2 + \dfrac{15}{16}\, U a^2. & (3.8) \end{cases}$$

We note that

$$v - u = 0,\ \psi_i - \varphi_i = 0 \text{ at } r = a. \tag{3.9}$$

All these results will be useful for later calculations.

4. *The correlation tensors due to Hill's vortices, all of one type*

Consider two points A_0 and $\tilde{A}_0$ in an isotropic homogeneous field of turbulence due to a distribution of Hill's vortices. On the whole, it is easier to use general Cartesian axes, rather than axes specially related to the line $\tilde{A}_0 A_0$. Any slight simplification arising from special axes is outweighed by the symmetry obtained from the use of general axes and indicial notation.

Let v_i, $\tilde{v}_k$ be the velocities of fluctuation at the points A_0, $\tilde{A}_0$, i.e., the fluid velocities after subtraction of the steady streaming, in any particular observation. The correlation tensors of von Kármán and Howarth (3) are defined in terms of the average values as follows:

$$\begin{cases} R_{ik} = \dfrac{v_i \tilde{v}_k}{v^2}, \quad T_{ikl} = \dfrac{v_i v_k \tilde{v}_l}{(v^2)^{3/2}}, \text{ etc.,} \\[2ex] v^2 = v_1^2 = v_2^2 = v_3^2 = \frac{1}{3} v_i v_i = \frac{1}{3} \tilde{v}_k \tilde{v}_k. \end{cases} \tag{4.1}$$

The physical meaning of R_{ik} is this: if λ_i, $\tilde{\lambda}_i$ are unit vectors at A_0, $\tilde{A}_0$ respectively, then the invariant $R_{ik}\lambda_i\tilde{\lambda}_k$ is the correlation coefficient between the components of the velocities at A_0 and $\tilde{A}_0$ in the directions of λ_i and $\tilde{\lambda}_i$.

We shall first calculate these correlation tensors for a distribution of Hill's vortices, all of one type (i.e., of given U and a), passing on in §5 to the more general case where there is a distribution in type.

Let us return to Fig. 1 and consider a single Hill's vortex pointing in the direction l_i. Denote the distance $\tilde{A}_0 A_0$ by ξ. We shall use the formulae of §3, with the notation

$$\begin{cases} r_i = x_i - x_i^0, \\ \tilde{r}_i = x_i - \tilde{x}_i^0. \end{cases} \tag{4.2}$$

In addition to the formulae written explicitly in §3, we have also those obtained on replacing x_i^0 by $\tilde{x}_i^0$, thus leading to expressions $\tilde{\alpha}_{ij}$, $\tilde{\beta}_{ij}$ and other quantities so marked. Thus we have, for the velocities at A_0 and $\tilde{A}_0$, due to the single vortex,

$$u_i = \alpha_{ij} l_j \quad (r \geq a), \qquad u_i = \beta_{ij} l_j \quad (r \leq a), \tag{4.3}$$

$$\tilde{u}_k = \tilde{\alpha}_{kj} l_j \quad (\tilde{r} \geq a), \qquad \tilde{u}_k = \tilde{\beta}_{kj} l_j \quad (\tilde{r} \leq a), \tag{4.4}$$

where α_{ij}, β_{ij} are as in (3.1a), (3.1b) and $\tilde{\alpha}_{kj}$, $\tilde{\beta}_{kj}$ are corresponding expressions with r, r_i replaced by $\tilde{r}$, $\tilde{r}_i$. The formulae to be applied depend on whether the points A_0, $\tilde{A}_0$ lie outside or inside the vortex. The symbols v_i, $\tilde{v}_k$ in (4.1) denote velocities due to the whole distribution of vortices in a single observation; u_i, $\tilde{u}_k$ in (4.3), (4.4) are due to a single typical vortex.

Before applying (2.13), we should be sure that (4.3) and (4.4) are in conformity with (2.8a), i.e.,

$$\bar{u}_i = 0, \qquad \bar{\tilde{u}}_k = 0. \tag{4.5}$$

This can be seen at once, from the assumed isotropy, which implies

$$\bar{l}_j^{\,l} = 0. \tag{4.6}$$

We can therefore use equations (2.13) to calculate the correlation tensors of the second order, (2.14) for the third order, and so on.

According to those equations, the correlation tensors are proportional to

$$\int \overline{u_i \tilde{u}_k}^{\,l}\, d\tau, \quad \int \overline{u_i u_k \tilde{u}_l}^{\,l}\, d\tau, \text{ etc.,} \tag{4.7}$$

the volume integrals being taken through the whole space. The first step is therefore to evaluate the average with respect to direction. Referring to (4.3) and (4.4), we see that this average is at once given with the help of the following formulae, which can be easily verified:

$$\begin{cases} \overline{l_i l_j}^{\,l} = \frac{1}{3}\delta_{ij}, \\ \overline{l_i l_j l_k}^{\,l} = 0, \\ \overline{l_i l_j l_k l_m}^{\,l} = \frac{1}{15}(\delta_{ij}\delta_{km} + \delta_{ik}\delta_{jm} + \delta_{im}\delta_{jk}), \text{ etc.} \end{cases} \tag{4.8}$$

Thus, all correlations of odd orders vanish, while those of even orders do not vanish in general. This is contrary to the view of Taylor (8), who has argued that the third order correlations must not vanish identically. However, this discrepancy may be due to an oversimplification in the choice of the model of an eddy. With a more elaborate model, the formulae (4.3) and (4.4) for u_i and $\tilde{u}_k$ might contain terms quadratic in the direction cosines, and then the correlations of odd orders would not necessarily vanish. The vanishing of the odd-order correlations may be attributed to the symmetry of the vortex, fore and aft. Our main interest, however, lies in correlations of the second order, because they are those which are actually measured.

Since the formal expressions (4.3) and (4.4) for the velocity are different for $r > a$ and $r < a$, it is evident that we must distinguish two cases in our calculations, according as the distance ξ between the points of observation is greater or less than the diameter $2a$ of the vortex.

Case (1) $\xi \geq 2a$. This will give results corresponding to the outer part of the correlation function. In this case, no single vortex of radius a can contain both the points at which the fluctuations are observed.

Case (2) $\xi \leq 2a$. This will give results corresponding to the inner part of the correlation functions. We shall sub-divide this case into two cases:

(2a) $a \leq \xi \leq 2a$,

(2b) $0 \leq \xi \leq a$.

In either of these cases a single vortex may contain both the points at which the fluctuations are observed.

In Fig. 3, we draw spheres of radius a around the points A_0 and $\tilde{A}_0$. Cases (1), (2a), and (2b) are shown separately. In Case (1) there are four regions I, $\tilde{I}$, II, $\tilde{II}$, bounded by spheres S, $\tilde{S}$ of radius a centred on A_0, $\tilde{A}_0$, and by the plane of symmetry Π. In the other cases there is a similar but more complicated division of space, as may be seen from the diagrams.

Now, if a vortex has its centre at the point A, the velocities at A_0, $\tilde{A}_0$ will depend of course on the position of A. But the formal expression for the velocity will also depend on whether the point of observation lies inside or outside the vortex. Hence, we have the following table (Table I), which shows the formulae for u_i, $\tilde{u}_k$, corresponding to the cases where A lies in the various regions.

Referring to Table I and equation (4.8), we see that the integrals to be evaluated are as follows.

Case (1):

$$(4.9) \quad \int \overline{u_i \tilde{u}_k}' \, d\tau = \tfrac{1}{3} \int_{I+\tilde{I}} \alpha_{ij}\, \tilde{\alpha}_{kj}\, d\tau + \tfrac{1}{3} \int_{II} \beta_{ij}\, {}^{\alpha}{}_{kj}\, d\tau + \tfrac{1}{3} \int_{\tilde{II}} \alpha_{ij}\, \tilde{\beta}_{kj}\, d\tau.$$

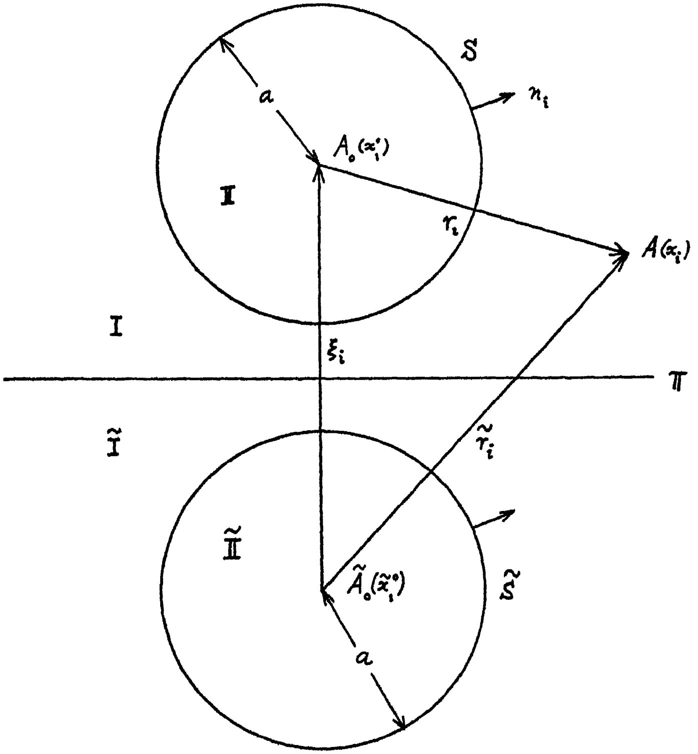


FIGURE 3—Case (1) $\xi \geq 2a$

FIGURE 3.—Partition of space regions for different cases, classified according to the ratio $\xi/2a$.

A = centre of vortex,

$\tilde{A}_0, A_0$ = points where the fluctuations are observed,

$\xi = \tilde{A}_0 A_0$ = distance between the points of observation,

a = radius of the Hill's vortex,

$\xi_i = x_i^0 - \tilde{x}_i^0, \qquad r_i = x_i - x_i^0, \qquad \tilde{r}_i = x_i - \tilde{x}_i^0.$

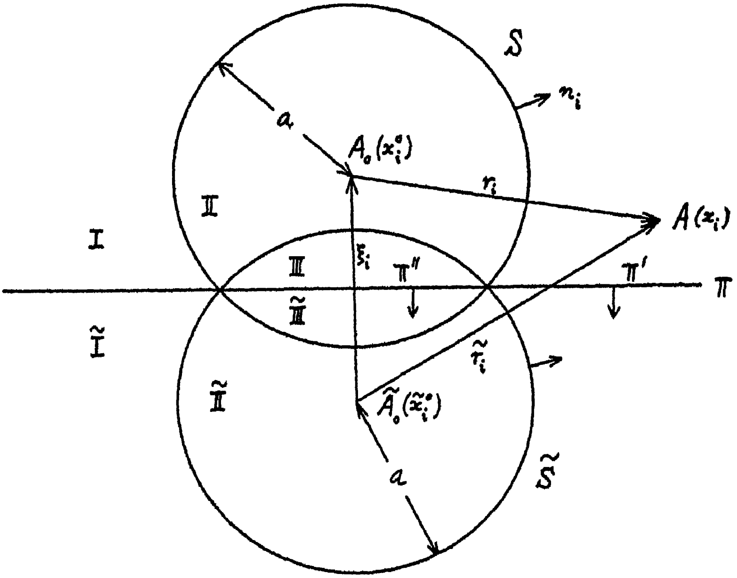


FIGURE 3—Case (2*a*) $a \leq \xi \leq 2a$

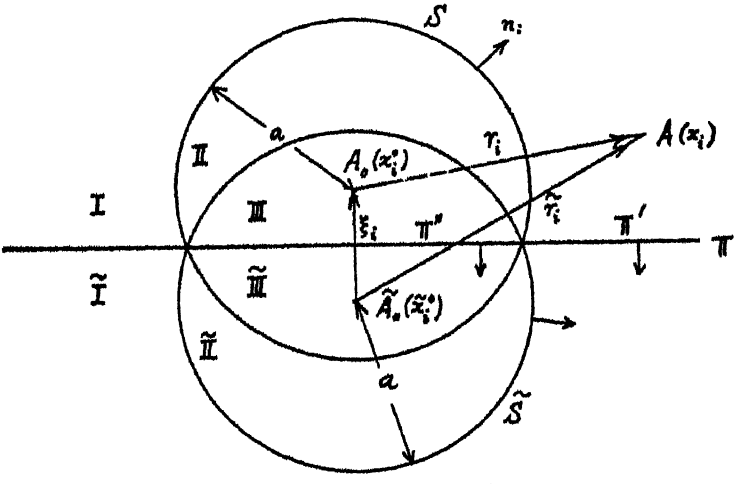


FIGURE 3—Case (2*b*) $0 \leq \xi \leq a$

Case (2):

$$(4.10)\quad \int \overline{u_i \tilde{u}_k}^{\,l}\, d\tau = \tfrac{1}{3}\int_{I} a_{ij}\,\tilde{a}_{kj}\,d\tau + \tfrac{1}{3}\int_{II} \beta_{ij}\,\tilde{a}_{kj}\,d\tau + \tfrac{1}{3}\int_{III} \beta_{ij}\,\tilde{\beta}_{kj}\,d\tau$$
$$+ \tfrac{1}{3}\int_{\tilde{I}} a_{ij}\,\tilde{a}_{kj}\,d\tau + \tfrac{1}{3}\int_{\widetilde{II}} a_{ij}\,\tilde{\beta}_{kj}\,d\tau + \tfrac{1}{3}\int_{\widetilde{III}} \beta_{ij}\,\tilde{\beta}_{kj}\,d\tau.$$

TABLE I

	Region containing the centre (A) of the vortex					
	I	II	III	$\tilde{I}$	$\widetilde{II}$	$\widetilde{III}$
u_i	$a_{ij}\, l_j$	$\beta_{ij}\, l_j$	$\beta_{ij}\, l_j$	$a_{ij}\, l_j$	$a_{ij}\, l_j$	$\beta_{ij}\, l_j$
$\tilde{u}_k$	$a_{kj}\, l_j$	$\tilde{a}_{kj}\, l_j$	$\tilde{\beta}_{kj}\, l_j$	$\tilde{a}_{kj}\, l_j$	$\tilde{\beta}_{kj}\, l_j$	$\tilde{\beta}_{kj}\, l_j$

The method of evaluation employed below consists in transforming the volume integrals into suitable surface integrals with the help of Green's theorem. Let us consider the cases separately.

Case (1). To evaluate (4.9) we first transform the integrals as follows. By (3.4) and (3.5), we have

$$a_{ij}\,\tilde{a}_{kj} = \frac{\partial \varphi_i}{\partial x_j}\,\tilde{a}_{kj} = \frac{\partial}{\partial x_j}(\varphi_i\,\tilde{a}_{kj}),$$

$$\beta_{ij}\,\tilde{a}_{kj} = \frac{\partial}{\partial x_j}(\beta_{ij}\,\tilde{\varphi}_k),$$

$$a_{ij}\,\tilde{\beta}_{kj} = \frac{\partial}{\partial x_j}(\tilde{\beta}_{kj}\,\varphi_i).$$

The integral $(I + \tilde{I})$ in (4.9) transforms by Green's theorem into surface integrals over S and $\tilde{S}$ and an infinite sphere, this last integral vanishing. The integrals over II and $\widetilde{II}$ transform into surface integrals over S and $\tilde{S}$ respectively, so that we are left with four surface integrals. Two of these cancel by (3.3), and we obtain finally

$$(4.11)\quad \int \overline{u_i\,\tilde{u}_k}^{\,l}\, d\tau = \tfrac{1}{3}\int_{S} (-\,\varphi_i\,\tilde{a}_{kj} + \tilde{\varphi}_k\beta_{ij})\,n_j\,dS,$$

where n_j is the unit normal vector to S, drawn outward. With a little manipulation and use of some well-known results in potential theory (see Appendix), the whole integral finally reduces to

$$(1.12)\quad \int u_i \bar{\tilde{u}}_k{}^l \, d\tau = \frac{\pi}{3}(Ua^3)^2 \frac{\partial^2}{\partial\xi_i \partial\xi_k}\left(\frac{1}{\xi}\right) = \frac{\pi}{3}(Ua^3)^2\left(-\frac{\delta_{ik}}{\xi^3} + \frac{3\xi_i\xi_k}{\xi^5}\right),$$

where

$$\xi_i = x_i^0 - \tilde{x}_i^0,$$

the coordinates of A_0 relative to $\tilde{A}_0$, and $\xi^2 = \xi_i\xi_i$.

Case (2). There are six integrals to be evaluated in (1.10). By symmetry, the last three integrals are equal to the first three, with the suffixes i, k interchanged. We may therefore write

$$(4.13)\quad \int u_i \tilde{u}_k{}^l \, d\tau = \tfrac{1}{2}(S_{ik} + S_{ki}),$$

where

$$(4.14)\quad S_{ik} = \int_I a_{ij}\tilde{a}_{kj}\,d\tau + \int_{II} \beta_{ij}\tilde{a}_{kj}\,d\tau + \int_{III} \beta_{ij}\tilde{\beta}_{kj}\,d\tau.$$

Unfortunately, the simple method used for Case (1) cannot be applied here. The first step in evaluating these integrals is to transform them by similar methods into surface integrals over the plane of symmetry,—a transformation which might also have been used in Case (1) to give the result (4.12).

Let Π' and Π'' denote respectively the portions of the plane of symmetry $r = \tilde{r}$ for which $r > a$ and $r < a$. We transform the first integral in (4.14) into integrals over Π' and a portion of S; we transform the second integral into integrals over portions of S and $\tilde{S}$. On addition, the integrals over S cancel by (3.3), and the integral over $\tilde{S}$ is then transformed into an integral over Π'' and a volume integral over III. Thus we obtain

$$(4.15)\quad S_{ik} = \int_{\Pi'} n_j a_{ij}\tilde{\varphi}_k\,dS + \int_{\Pi''} n_j \beta_{ij}\tilde{\varphi}_k\,dS + \int_{III} (\tilde{\beta}_{kj} - \tilde{a}_{kj})\beta_{ij}\,d\tau,$$

where n_j is the unit normal to Π' and Π'', pointing away from x_i^0, as indicated in Fig. 3.

Substituting for β_{ij} from (3.6) in the volume integral, and using (3.7), (3.8), (3.9), we may reduce it in part to a surface integral over Π'' as follows:

$$(4.16)\quad \int_{III} (\tilde{\beta}_{kj} - \tilde{a}_{kj})\beta_{ij}\,d\tau = \int_{\Pi''} \{n_j\psi_i(\tilde{\beta}_{kj} - \tilde{a}_{kj}) + n_i(\tilde{\psi}_k - \tilde{\varphi}_k)q$$
$$- n_k(\tilde{v} - \tilde{u})q'' r_i\}\,dS + \delta_{ik}\int_{III} \{q''(\tilde{v} - \tilde{u}) + q\tilde{q}\}\,d\tau,$$

where
$$q'' = \frac{d^2q}{dr^2} = -\frac{15}{2}\frac{U}{a^2}.$$

By using bipolar coordinates $(r, \tilde{r}, \varphi)$, where φ is the azimuthal angle about the line $\tilde{A}_0A_0$, we can show (see Appendix) that the surface integrals in (4.15) and (4.16) are given by

$$(4.17)\qquad \int_{\text{II}'} n_j\alpha_{ij}\tilde{\varphi}_k\,dS = \frac{\pi U^2a^3}{32}\{(-6\eta + 4\eta^3)\delta_{ik} + (2\eta^{-1} + 4\eta)\eta_i\eta_k\},$$

$$(4.18)\qquad \int_{\text{II}''} n_j\beta_{ij}\tilde{\varphi}_k\,dS = \frac{\pi U^2a^3}{32}\{(-24\eta + 48\eta^2 - 24\eta^3)\delta_{ik} + (80\eta^{-2} - 152\eta^{-1} + 96 - 24\eta)\eta_i\eta_k\},$$

$$(4.19)\qquad \int_{\text{II}''}\{n_j\psi_i(\tilde{\beta}_{kj} - \tilde{\alpha}_{kj}) + n_i(\tilde{\psi}_k - \tilde{\varphi}_k)q - n_k(\tilde{v} - \tilde{u})q''r_i\}\,dS$$
$$= \frac{\pi U^2a^3}{32}\{(-40 + 105\eta - 72\eta^2 - 5\eta^3 + 15\eta^5 - 3\eta^7)\delta_{ik}$$
$$+(-80\eta^{-2} + 150\eta^{-1} + 96 - 280\eta + 150\eta^3 - 36\eta^5)\eta_i\eta_k\},$$

where

$$(4.20)\qquad \eta_i = \xi_i/2a,\quad \eta = \xi/2a.$$

The volume integral in (4.16) (for both cases (2a) and (2b)) is

$$(4.21)\qquad \delta_{ik}\int_{III}\{q''(\tilde{v} - \tilde{u}) + q\tilde{q}\}\,d\tau = \frac{\pi U^2a^3}{32}\left\{\frac{600}{7} - 75\eta - 360\eta^2 + 525\eta^3 - 225\eta^5 + \frac{345}{7}\eta^7\right\}\delta_{ik}.$$

The quantities S_{ik} are evaluated by adding together (4.17), (4.18), (4.19) and (4.21), and so by (4.13) we have for Case (2)

$$(4.22)\qquad \int\overline{u_i\vec{u}_k}^{\,l}\,d\tau = \frac{\pi U^2a^3}{48}\left\{\left(\frac{320}{7} - 384\eta^2 + 500\eta^3 - 210\eta^5 + \frac{324}{7}\eta^7\right)\delta_{ik} + (192\eta^2 - 300\eta^3 + 150\eta^5 - 36\eta^7)\frac{\eta_i\eta_k}{\eta^2}\right\}.$$

The above calculations involve vortices of one type (U, a). If we suppose that there is actually only one type present, our calculation of correlation functions is nearly completed. If on the other hand, we suppose that there is a distribution in type, what we have done is only an essential part in the whole calculation, for averaging with respect to type has still to be carried out.

In order to check our results, we shall for the present assume that only one type of vortex is present. The case of a distribution in type will be treated in §5.

For the correlation tensor R_{ik} we have, as in (4.1),

$$R_{ik} = \frac{\overline{v_i \tilde{v}_k}}{v^2}, \quad v^2 = \tfrac{1}{3} v_j v_j.$$

Here v_i, $\tilde{v}_i$ are velocities at A_0, $\tilde{A}_0$ respectively, corresponding to a single observation. Hence, by (2.13), since there is only one type of vortex present,

$$\text{(4.23)} \qquad \overline{v_i \tilde{v}_k} = \rho \int u_i \tilde{u}_k^{\,l}\, d\tau.$$

We obtain v^2 by putting $k = i$ (no summation) and letting $\xi \to 0$ in (4.22); thus

$$\text{(4.23a)} \qquad v^2 = \frac{1}{3}\, \overline{v_i v_i} = \rho \frac{\pi U^2 a^3}{48} \cdot \frac{320}{7}.$$

Let us write (4.12) in the form

$$\text{(4.24)} \qquad \int u_i \tilde{u}_k^{\,l}\, d\tau = \frac{\pi U^2 a^3}{48} \left(-\frac{2\delta_{ik}}{\eta^3} + \frac{6\eta_i \eta_k}{\eta^5} \right).$$

Hence, from this equation and (4.22) and (4.23a), the correlation tensor may be written

$$\text{(4.25)} \qquad R_{ik} = G(\eta)\delta_{ik} + \left(F(\eta) - G(\eta) \right) \frac{\eta_i \eta_k}{\eta^2}, \quad \eta_i = \frac{\xi_i}{2a},$$

where the functions F, G take different forms according as $\eta >$ or < 1:

$$\text{(4.26a)} \qquad \left\{ \begin{aligned} F(\eta) &= \frac{7}{80} \frac{1}{\eta^3}, \\ G(\eta) &= -\frac{7}{160} \frac{1}{\eta^3}, \end{aligned} \right\} \quad (\eta > 1);$$

$$\text{(4.26b)} \qquad \left\{ \begin{aligned} F(\eta) &= 1 - \frac{21}{5}\eta^2 + \frac{35}{8}\eta^3 - \frac{21}{16}\eta^5 + \frac{9}{40}\eta^7, \\ G(\eta) &= 1 - \frac{42}{5}\eta^2 + \frac{175}{16}\eta^3 - \frac{147}{32}\eta^5 + \frac{81}{80}\eta^7, \end{aligned} \right\} \quad (\eta < 1).$$

We may also write (4.25) as

$$\text{(4.27)} \qquad R_{ik} = g(\xi)\delta_{ik} + [f(\xi) - g(\xi)] \frac{\xi_i \xi_k}{\xi^2},$$

where

$$(4.28)\qquad \begin{cases} f(\xi) = F(\eta) = F\left(\dfrac{\xi}{2a}\right), \\ g(\xi) = G(\eta) = G\left(\dfrac{\xi}{2a}\right). \end{cases}$$

The correlation tensor is thus expressed in a usual form. Indeed, from the mere assumption of isotropy (without any particular microscopic model) we know* that R_{ik} must be of the general form (4.27).

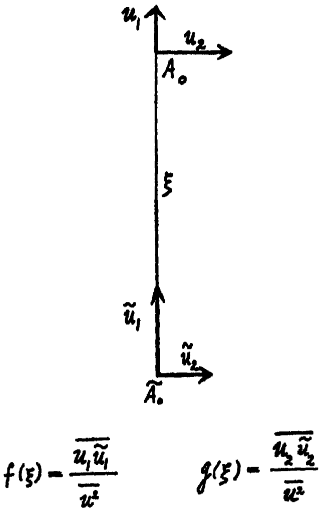

u_1
u_2
A_0
ξ
$\tilde{u}_1$
$\tilde{u}_2$
$\tilde{A}_0$
$f(\xi) = \dfrac{\overline{u_1\tilde{u}_1}}{\overline{u^2}}$
$g(\xi) = \dfrac{\overline{u_2\tilde{u}_2}}{\overline{u^2}}$

FIGURE 4.—Longitudinal (f) and transverse (g) correlation functions ($\tilde{A}_0$, A_0 being the points of observation).

The function $f(\xi)$ and $g(\xi)$ are the longitudinal and transverse correlation functions respectively (Fig. 4). In accordance with general theory, they must satisfy (as a mere deduction from isotropy and the equation of continuity) von Kármán's equation

$$(4.29)\qquad f - g = -\tfrac{1}{2}\xi\frac{df}{d\xi}.$$

We easily verify that the above functions do in fact satisfy this relation. This is a useful check on the accuracy of the complicated calculations involved.

---

*Cf. eq. (3), p. 195 of the paper by von Kármán and Howarth, *loc. cit.*; see also Robertson, *loc. cit.*

Let us sum up our result: *On the assumption that a state of isotropic and homogeneous turbulence is due to Hill's spherical vortices all of one speed U and one radius a, behaving independently, the longitudinal (f) and transverse (g) correlation functions for points at a distance ξ apart are as follows*:

$$\text{(4.30a)} \qquad \left\{\begin{aligned} f(\xi) &= \frac{7}{80}\left(\frac{2a}{\xi}\right)^3 \\ g(\xi) &= -\frac{7}{160}\left(\frac{2a}{\xi}\right)^3 \end{aligned}\right| \quad \text{for } \xi > 2a,$$

(4.30b)

$$\left\{\begin{aligned} f(\xi) &= 1 - \frac{21}{5}\left(\frac{\xi}{2a}\right)^2 + \frac{35}{8}\left(\frac{\xi}{2a}\right)^4 - \frac{21}{16}\left(\frac{\xi}{2a}\right)^5 + \frac{9}{40}\left(\frac{\xi}{2a}\right)^7 \\ g(\xi) &= 1 - \frac{42}{5}\left(\frac{\xi}{2a}\right)^2 + \frac{175}{32}\left(\frac{\xi}{2a}\right)^3 - \frac{147}{32}\left(\frac{\xi}{2a}\right)^5 + \frac{81}{80}\left(\frac{\xi}{2a}\right)^7 \end{aligned}\right| \quad \text{for } \xi < 2a.$$

We would, of course, expect the correlation functions to join up continuously at $\xi = 2a$; but the junction is remarkably smooth, there being continuity of derivatives up to the fourth order inclusive for $f(\xi)$ and up to the third order inclusive for $g(\xi)$.

We note that the correlation functions are independent of U, the speed of the vortices. Their dependence on a is very simple, a change in a corresponding to a change of scale.

Before passing to the consideration of distribution in type, let us connect our model with the micro- and macro-scales of turbulence.

The micro-scale of turbulence defined by Taylor (2) is given by

$$\text{(4.31)} \qquad -\frac{1}{\lambda^2} = \left(\frac{d^2 f}{d\xi^2}\right)_{\xi=0},$$

while the macro-scale is given by

$$\text{(4.32)} \qquad l = \int_0^\infty g(\xi)\,d\xi$$

or, by (4.29), provided that $\xi f(\xi)$ vanishes for infinite ξ (as it does in our theory),

$$\text{(4.33)} \qquad l = \tfrac{1}{2}\int_0^\infty f(\xi)\,d\xi.$$

For distinction we shall use capitals (Λ and L) for the case where there is only one type of vortex present.

With the function $f(\xi) = F(\eta)$ given by (4.26a) and (4.26b), we find that the micro-scale Λ is given by

(4.34)
$$\frac{1}{\Lambda^2} = -\frac{1}{4a^2}F''(0) = \frac{21}{10}\frac{1}{a^2},$$

or

(4.35)
$$\Lambda = \sqrt{\frac{10}{21}}\,a = 0.69\,a,$$

and the macro-scale L is given by

(4.36)
$$L = a\int_0^\infty F(\eta)\,d\eta = \frac{35}{64}a = 0.55\,a.$$

We observe that the lengths L and Λ are directly connected with the size of the eddies, Λ being about two-thirds of the radius of an eddy.

Comparison of the shape of the correlation curve given by (4.26a) and (4.26b) with other results (theoretical and experimental) will be made in §6.

5. *The correlation tensors due to Hill's vortices with distribution in type*

So far we have dealt only with vortices of one type, all of the same size and travelling with the same speed. To treat the general case, we have to include a further averaging process; we must integrate with respect to the variables U and a in (2.13) and (2.14). We note that the correlation tensors of the third order are still identically zero. For the correlations of the second order, we now have, instead of (4.23),

(5.1)
$$\overline{v_i\tilde{v}_k} = \rho\int\Phi(U,a)\,dUda\int\overline{u_i\tilde{u}_k}^{\,l}\,d\tau;$$

for the last integral we have the expressions (4.22) or (4.24) according as $2a$ is greater or less than ξ. Also

(5.1a)
$$\overline{v_i v_i} = \rho\int\Phi(U,a)\,dUda\lim_{\xi\to 0}\int\overline{u_i\tilde{u}_i}^{\,l}\,d\tau.$$

For the limit, we have again the expression (4.23a). It then follows that the correlation tensor is

5.2)
$$\mathfrak{R}_{ik} = \frac{\displaystyle\int_0^\infty\int_0^\infty \Phi(U,a)\,U^2a^3\left\{G\left(\frac{\xi}{2a}\right)\delta_{ik} + \left[F\left(\frac{\xi}{2a}\right) - G\left(\frac{\xi}{2a}\right)\right]\frac{\xi_i\xi_k}{\xi^2}\right\}dUda}{\displaystyle\int_0^\infty\int_0^\infty \Phi(U,a)\,U^2a^3\,dUda},$$

where the functions F and G are as in (4.26a) if $2a < \xi$ and as in (4.26b) if $2a > \xi$. If we write R_{ik} in the standard form

$$R_{ik} = g(\xi)\delta_{ik} + [f(\xi) - g(\xi)]\frac{\xi_i \xi_k}{\xi^2}, \tag{5.3}$$

we obtain the following expressions for the longitudinal and transverse correlation functions:

$$f(\xi) = \int_0^\infty H(a) F\left(\frac{\xi}{2a}\right) da, \tag{5.1}$$

$$g(\xi) = \int_0^\infty H(a) G\left(\frac{\xi}{2a}\right) da, \tag{5.5}$$

where

$$H(a) = \frac{\displaystyle\int_0^\infty \Phi(U,a)\, U^2 a^3 dU}{\displaystyle\int_0^\infty \int_0^\infty \Phi(U,a)\, U^2 a^3 dU da}, \tag{5.6}$$

we note that

$$\int_0^\infty H(a) da = 1. \tag{5.7}$$

Since the energy associated with a Hill's vortex is

$$\frac{10}{7}\pi\rho U^2 a^3,$$

it follows from (5.6) that $H(a)$ may be interpreted as giving the distribution of energy among the sizes of the vortices. Our microscopic model gives no information at all regarding this function; we can only make some simple and plausible assumption.

Let us assume

$$H(a) = A a^2 e^{-a/a_0}, \tag{5.8}$$

where A and a_0 are constants. These constants are connected by (5.7), and so we get

$$H(a) = \frac{1}{2a_0}\left(\frac{a}{a_0}\right)^2 e^{-a/a_0}, \tag{5.9}$$

the constant a_0 remaining undetermined. With this choice of $H(a)$, the correlation functions (5.4) and (5.5) become (with $v = a/a_0$)

$$f(\xi) = \int_0^\infty \tfrac{1}{2} v^2 e^{-v} F\left(\frac{\xi}{2a_0 v}\right) dv, \tag{5.10}$$

$$g(\xi) = \int_0^\infty \tfrac{1}{2} v^2 e^{-v} G\left(\frac{\xi}{2a_0 v}\right) dv. \tag{5.11}$$

When (4.26a) and (4.26b) are substituted into these formulae, a direct calculation gives

$$\begin{aligned} f(\xi) = {} & \frac{7}{160}\zeta^{-3}\{120 - e^{-\zeta}(\zeta^5 + 5\zeta^4 + 20\zeta^3 + 60\zeta^2 + 120\zeta + 120)\} \\ & + e^{-\zeta}\left\{1 + \zeta - \frac{16}{10}\zeta^2 - \frac{3}{10}\zeta^3 + \frac{51}{160}\zeta^4 + \frac{3}{640}\zeta^5 - \frac{3}{640}\zeta^6\right\} \\ & + \int_\zeta^\infty \frac{e^{-v}}{v}\,dv\left\{\frac{35}{16}\zeta^3 - \frac{21}{64}\zeta^5 + \frac{3}{640}\zeta^7\right\}, \end{aligned} \tag{5.12}$$

(5.13)

$$\begin{aligned} g(\xi) = {} & -\frac{7}{320}\zeta^{-3}\{120 - e^{-\zeta}(\zeta^5 + 5\zeta^4 + 20\zeta^3 + 60\zeta^2 + 120\zeta + 120)\} \\ & + e^{-\zeta}\left\{1 + \zeta - \frac{37}{10}\zeta^2 - \frac{327}{320}\zeta^3 + \frac{177}{160}\zeta^4 + \frac{27}{1280}\zeta^5 - \frac{27}{1280}\zeta^6\right\} \\ & + \int_\zeta^\infty \frac{e^{-v}}{v}\,dv\left\{\frac{175}{32}\zeta^3 - \frac{147}{128}\zeta^5 + \frac{27}{1280}\zeta^7\right\}, \end{aligned}$$

where

$$\zeta = \xi/2a_0. \tag{5.14}$$

The discontinuities in the higher derivatives of $f(\xi)$ and $g(\xi)$ have been removed by the process of averaging for type; the functions are now analytic. As a check on the calculation, we may verify that they still satisfy the relation (4.29).

Referring to equations (4.31) – (4.36) and (5.4), we see that the micro-scale of turbulence λ is given by

$$\frac{1}{\lambda^2} = \int_0^\infty \frac{1}{\Lambda^2} H(a)da = \frac{21}{10}\int_0^\infty \frac{1}{a^2} H(a)da, \tag{5.15}$$

and that the macro-scale l is given by

$$l = \int_0^\infty LH(a)da = \frac{35}{64}\int_0^\infty H(a)a\,da. \tag{5.16}$$

With $H(a)$ given by (5.9), we obtain

$$\frac{1}{\lambda^2} = \frac{21}{20}\frac{1}{a_0^2}, \tag{5.17}$$

or

$$\lambda = \sqrt{\frac{20}{21}}\,a_0 = 0.98\,a_0, \tag{5.18}$$

and

$$l = \frac{105}{64} a_0 = 1.64\, a_0. \tag{5.19}$$

The distribution of energy presents a maximum (given by $H'(a) = 0$) at

$$a = 2a_0. \tag{5.20}$$

We see that λ is about one-half of the radius of eddy corresponding to maximum share of energy.

To sum up: *We consider a state of isotropic and homogeneous turbulence corresponding to Hill's vortices with a distribution in type* (U, a) *such that the function* $H(a)$, *defined by (5.6) in terms of the probability* $\Phi(U, a)dUda$ *of a vortex being in the range* $(U, U + dU; a, a + da)$, *is given by (5.9). Then the longitudinal* (f) *and transverse* (g) *correlation functions for points at a distance* ξ *apart are given by (5.12) and (5.13), together with (5.14) where* a_0 *is the constant appearing in (5.9). The micro-scale* (λ) *of turbulence defined by (4.31) and the macro-scale* (l) *defined by (4.32) are given by (5.18) and (5.19) respectively.*

Comparison with experiments will be made in the next section.

6. *Comparison with experiment*

Let us now try to compare with experimental data the theoretical correlation functions $f(\xi)$ and $g(\xi)$, as given by (5.12) and (5.13), as well as the simpler functions given by (4.26a), (4.26b) and (4.28). Since the relation (4.29) (which is also experimentally verified (8), (9)) gives the function $g(\xi)$ explicitly in terms of $f(\xi)$, we shall discuss the function $f(\xi)$ alone.

Before comparing the exact shape of the theoretical and experimental correlation curves, let us first consider the question of "similarity."

If we consider all the vortices to be of one type (§ 1), the correlation function $f(\xi)$ is given by the first equation of (4.28). Considered as a function of $\eta (= \xi/2a)$, it has a definite form. It is, however, more natural from a physical point of view to consider it as a function of ξ, the distance between the points of observation. If, then, we compare correlation functions at two places, one at some distance down-stream from the other, we are to expect that, due to decay, the vortices at the two places will be of different types, and the function $f(\xi)$ will therefore have different forms at the two places. But the curves at the two places will be "similar," in the sense that the curve at one place may be obtained from the one at the other by a magnification in the direction of the ξ-axis. This means that the function $f(\xi)$ has a "self-preserving" property (3). Indeed, the correlation functions for different homogeneous and isotropic fields of turbulence, each due to vortices of a single type, will all be similar in the above sense. Referring to (5.4) and (5.12), we see clearly that this property of similarity still holds when we consider a distribution of vortices in type, if the function $H(a)$ as given by (5.9) is a universal one. (The arbitrary constant a_0 may vary from one case to the other.)

Let us now compare the theoretical results (5.12) for $f(\xi)$ with experimental data. We have the parameter a_0 in (5.12) to adjust for a change of scale of ξ. We take such a value for a_0 that the theoretical and experimental results agree for $f = 0.5$. The comparison is shown in Fig. 5, with $\zeta = \xi/2a_0$ as the horizontal coordinate. The experimental data are taken from Fig. 1 of Taylor's paper (8), reduced according to the relation $\zeta = 1.87\xi$ (ξ in inches). The agreement is good for the main part of the curve.*

The theoretical curve of $f(\xi)$ given by (4.26a) and (4.26b) is also shown for comparison. It is made by taking $\zeta = 2.45\eta$. This makes the curves agree for $f(\xi) = 0.5$. It can be seen that this curve is quite different from the other ones. Thus, the turbulence field behind a grid is not the one composed of Hill's vortices all of the same type.

Von Kármán and Howarth (3) have also derived the correlation curve of $f(\xi)$ for the case of "small Reynolds number of turbulence," on the assumption that the correlation functions are strictly "self-preserving" as we pass down-stream. Their results may be expressed as†

$$f(\xi) = \mathfrak{F}(\chi), \qquad \chi = \xi/b, \tag{6.1}$$

where b is an arbitrary constant, and where the function $\mathfrak{F}$ is uniquely defined when $\mathfrak{F}''(0)$ is given. Thus, for $\mathfrak{F}''(0) = -1/4$, $\mathfrak{F}(\chi) = e^{-\chi^2/8}$. This function is plotted in Fig. 5 by taking $\zeta = 0.488\chi$ so that it agrees with the other curves at $f(\xi) = 0.5$. It can be seen that their curve is similar to our curve of $f(\xi)$ given by (4.26a) and (4.26b) and bears only a slight resemblance to the experimental results. In Fig. 3 of their paper, the curves of $\mathfrak{F}(\chi)$ for $\mathfrak{F}''(0) = -1/4$ and $-1/5$ are shown. The latter curve when plotted according to the above scheme is nearly the the same as the former one, and consequently is not shown to avoid confusion.

---

*A question regarding the shape of the experimental curve in the immediate neighbourhood of $\xi = 0$, i.e., $\zeta = 0$, will be discussed below.

†Our ξ is their r. Our $\mathfrak{F}(\chi)$ is their $f(\chi)$, and is given by their equation (66) (in which $a = -\mathfrak{F}''(0)$). According to their notation, $\chi = r/\sqrt{\nu t}$, $\lambda^2 = \nu t/a^2$ and hence $\chi = r/b$, where $b = \lambda\sqrt{a}$, —a quantity varying with λ.

Let us now consider the question of similarity of the correlation curves in relation to experimental data. In the first place, experimental results do not seem to be definite enough to settle the question of "self-preservation" as we pass down-stream in a wind tunnel (10). Secondly, when we consider the correlation functions produced under

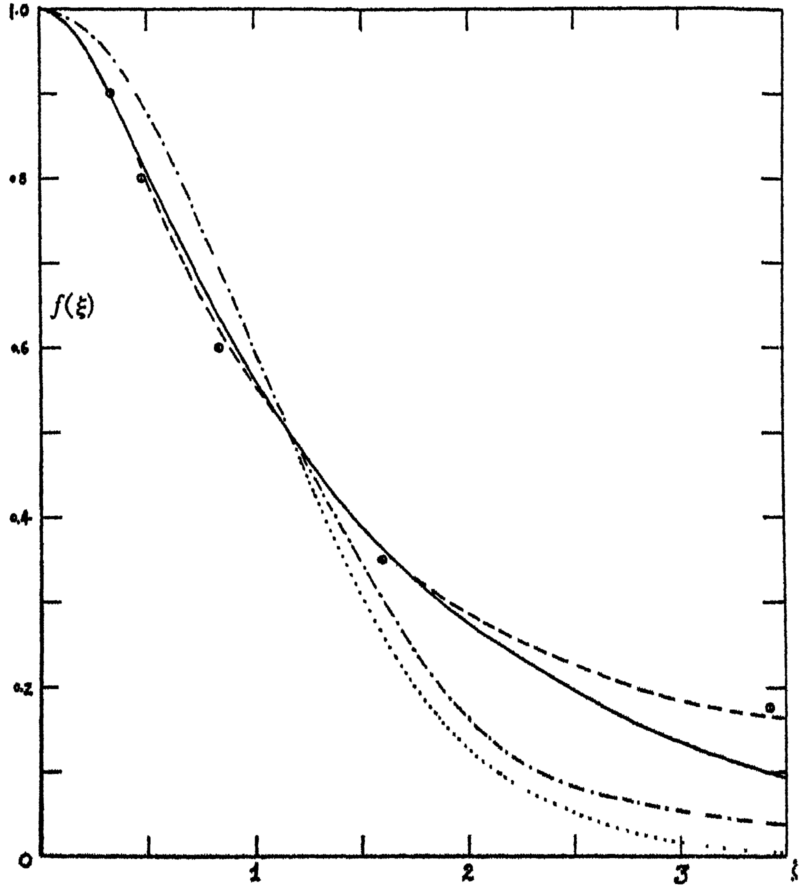


FIGURE 5.—Comparison of the theoretical correlation curves for $f(\xi)$ with experiment.

⊙ - - - - - - Experimental points and curve taken from Fig. 1 of Taylor's paper (8) with horizontal scale properly reduced ($\zeta = 1.87\xi$).

—————— Theoretical curve according to the present theory, vortices of different types.

— · — · — Theoretical curve according to the present theory, all vortices of the same type. Horizontal scale properly reduced ($\zeta = 2.45\eta$).

.... Theoretical curve according to von Kármán and Howarth for low Reynolds number of turbulence, $\mathfrak{F}''(0) = -1/4$. Horizontal scale properly reduced ($\zeta = 0.488\chi$).

different conditions, we have the following result of G. I. Taylor.* He wrote: "When the $R_1(r)$ curve is measured at a fixed section with varying wind-tunnel speeds, it is found to be practically invariable, except that the curvature at $r = 0$ increases as the wind speed increases." Thus, in the neighbourhood of $\xi = 0$, our theoretical result (5.12) can be in agreement with experiment only for a particular wind speed, for there the experimental curves differ among themselves. But for the main body of the curve, the agreement is good for all wind speeds (Fig. 5). This shows that our distribution function $H(a)$ as given by (5.9) (upon which (5.12) is based) is good for all wind speeds, except for slight variations with them.

Let us now go one step further and ask what kind of a variation in the distribution function $H(a)$ will give a change in the curve in the neighbourhood of $\xi = 0$, without changing the rest of the curve significantly. This point is illustrated by the following example.

Suppose that instead of the distribution function (5.9), we have

$$H(a) = \frac{1}{2a_0\{1+\frac{1}{2}\epsilon^k\Gamma(3-k)\}}\left\{1+\left(\frac{\epsilon}{v}\right)^k\right\}v^2e^{-v}, \quad v = \frac{a}{a_0}, \tag{6.2}$$

where Γ is the gamma-function

$$\Gamma(a) = \int_0^\infty v^{a-1}e^{-v}dv, \tag{6.2a}$$

and ϵ and k are positive constants, depending upon the experimental conditions under which the turbulence is produced, and satisfying the conditions that (i) ϵ is very small and that (ii) k is less than or equal to unity, but greater than a finite number, say 1/2. The two quantities ϵ and k may be related or not among themselves. The difference between (5.9) and (6.2) is very small and hence these two distribution functions give correlation curves not very much different from one another. But when we come to the calculation of the curvature at $\xi = 0$, we may find a considerable difference. By (5.15) and (6.2), we find that the curvature $1/\lambda'^2$ corresponding to (6.2) is given by

$$\frac{1}{\lambda'^2} = \frac{1}{\lambda^2}\,\frac{1+\epsilon^k\Gamma(1-k)}{1+\frac{1}{2}\epsilon^k\Gamma(3-k)}, \tag{6.3}$$

where λ is the micro-scale of turbulence according to (5.9) and is given by (5.17) or (5.18). When the two quantities ϵ and k vary, the value of λ' is changed. This means that the curvature of the correlation

---

*Taylor, *loc. cit.* (8), p. 315. His r is our ξ, and his $R_1(r)$ is our $f(\xi)$. In fact, he showed from his theory and experiment that the curvature at $\xi = 0$ varies directly as the wind speed.

curve $f(\xi)$ (say) at $\xi = 0$ changes with experimental conditions. In particular, if ϵ is zero, we have $\lambda' = \lambda$. If $k = 1$ and ϵ is not zero, (6.3) gives an infinite curvature for the curve at $\xi = 0$; this means that the curve has a pointed top. (This corresponds to an infinite rate of dissipation in the usual theory.) The curvature may be finite only when $H(a)$ vanishes at $a = 0$ faster than $O(a)$ (cf. eq. (5.15)). In a general manner, we may say that the correlation functions have sharper tops when small eddies are more abundant.

When ϵ and k are varying properly with the wind speed, the curvature given by (6.3) will be proportional to it. This serves to explain in general terms the above-mentioned experimental results of Taylor at various wind speeds. Since the correlation functions are nearly the same, the distribution function $H(a)$ must be nearly independent of the wind speed. The fact that the correlation curves are sharper at $\xi = 0$ for higher wind speeds means that there are then more small eddies. Referring to (5.15) and the example just mentioned, we see that a little change of $H(a)$ in the immediate neighbourhood of $a = 0$ is sufficient to bring about a big change in $1/\lambda^2$. We can therefore see why the correlation functions at different wind speeds can be very nearly the same as each other and yet have such a wide variation of curvature at the top. (This discussion is analogous to Taylor's discussion with the spectrum of turbulence (11).)

7. *Concluding remarks*

The fair agreement of the theoretical results with experiment seems to indicate the general correctness of this model of turbulence. In fact, it seems that no other calculation of correlation functions has led to results in fair agreement with experiment. The existing results regarding the shape of the correlation functions are von Kármán's relation (4.29) and Taylor's relation between frequency spectrum (11) and correlation functions. They are relations among observed quantities, but do not give theoretically the shape of the correlation functions themselves. The purely theoretical results derived by von Kármán and Howarth (8) regarding the correlation functions are true only for "low Reynolds number of turbulence," and it seems that experimental results satisfying this condition are still wanting.

The present theory also offers a very definite picture to interpret the scales of turbulence l and λ, which usually are only vaguely associated with the sizes of the eddies.

On the other hand, this theory also has its unsatisfactory points:

1. In the first place, Hill's spherical vortex is by no means a perfect model of an eddy. As mentioned above, the discontinuity of the

derivative of velocity across its surface makes its existence impossible in a viscous fluid. Owing to this discontinuity, it is difficult to build even an approximate theory of the decay of turbulence on the basis of the present model.

2. The correlation functions of the third order vanish identically in our model. This may be avoided by a small change in the model of an eddy, destroying the fore-and-aft symmetry of the Hill's vortex.

3. The distribution of energy with respect to size, or more generally, the distribution of Hill's vortices with respect to type, has not been treated satisfactorily. As discussed in §6, the distribution must be expected to be in general different for different fields. This lack of similarity is perhaps due to the effect of viscosity.

So far the theory deals only with the kinematics of homogeneous and isotropic turbulence. The next job in this theory is evidently to find the law of decay. As this is difficult with a model eddy involving discontinuity in the derivatives of velocity, we must try to find a reasonable model which is free from this defect. The new model should also be free from the defect of requiring all the correlation functions of odd order to vanish. So far the authors have not succeeded in this respect. The investigation of the nature of the distribution function, and particularly the effect of viscosity upon this function, is also a matter requiring careful investigation.

APPENDIX

We shall explain here briefly the calculations leading to the results (4.12) and (4.17)—(4.21). Some details are omitted, but sufficient detail is given to permit a check on the calculations.

Case (1). In this case, we are required to evaluate the integral given in (4.11). It is taken over a sphere, and so we can use the well-known Gauss theorem of mean value for harmonic functions. We deduce, in the notation of Fig. 3,

$$(1)\qquad \int_S \frac{n_i}{\tilde{r}}\, dS = \frac{4\pi}{3} a^3 \frac{\partial}{\partial \xi_i}\left(\frac{1}{\xi}\right);$$

this follows from the facts that

$$\frac{n_i}{\tilde{r}}\, dS = \int_{II} \frac{\partial}{\partial x_i}\left(\frac{1}{\tilde{r}}\right) d\tau,$$

$\frac{\partial}{\partial x_i}\left(\frac{1}{\tilde{r}}\right)$ is a harmonic function, and $\frac{\partial}{\partial x_i}\left(\frac{1}{\tilde{r}}\right) = \frac{\partial}{\partial \xi_i}\left(\frac{1}{\xi}\right)$ when $x_i = x_i^0$.

Let us now consider the first part of the integral in (4.11):

$$I_1 = -\int_S \phi_i \,\tilde{a}_{kj} n_j \, dS. \tag{2}$$

By (3.5) and (3.7), we can transform I_1 into

$$I_1 = -\frac{\partial^2}{\partial \tilde{x}_k^0 \partial \tilde{x}_j^0} \int_S \varphi_i \tilde{u} \, n_j \, dS. \tag{2a}$$

Writing out the integral $\int_S \varphi_i \tilde{u} \, n_j \, dS$ explicitly, noting that $r = a$ over S, and replacing r_i by $\tilde{r}_i - \xi_i$, we have

$$\int_S \varphi_i \tilde{u} \, n_j \, dS = \tfrac{1}{4} U^2 a^3 \int_S (\xi_i - \tilde{r}_i) \frac{1}{\tilde{r}} \, n_j \, dS. \tag{3}$$

Using (1) and Green's theorem, we can easily transform (3) into

$$\int_S \varphi_i \tilde{u} \, n_j \, dS = \tfrac{1}{4} U^2 a^3 \left\{ \frac{4\pi}{3} a^3 \xi_i \frac{\partial}{\partial \xi_j} \left(\frac{1}{\xi} \right) - \int_\Pi \frac{\partial^2}{\partial \tilde{x}_i^0 \partial \tilde{x}_j^0} (\tilde{r}) \, d\tau \right\}. \tag{4}$$

Substituting (4) in (2a) and noting that

$$\frac{\partial^2}{\partial \xi_j \partial \xi_j} \left(\frac{1}{\xi} \right) = 0, \quad \frac{\partial^2}{\partial \tilde{x}_j^0 \partial \tilde{x}_j^0} (\tilde{r}) = \frac{2}{\tilde{r}},$$

we have

$$I_1 = -\tfrac{1}{4} U^2 a^3 \left\{ \frac{4\pi}{3} a^3 \frac{\partial^2}{\partial \xi_i \partial \xi_k} \left(\frac{1}{\xi} \right) - \frac{\partial^2}{\partial \tilde{x}_i^0 \partial \tilde{x}_k^0} \int_\Pi \frac{2}{\tilde{r}} \, d\tau \right\}.$$

This gives

$$I_1 = \frac{\pi}{3} U^2 a^6 \frac{\partial^2}{\partial \xi_i \partial \xi_k} \left(\frac{1}{\xi} \right), \tag{5}$$

on applying Gauss' theorem.

The other part of the integral in (4.11)

$$I_2 = \int_S \tilde{\varphi}_k \beta_{ij} n_j \, dS \tag{6}$$

can be more easily evaluated. We have, by (3.7),

$$I_2 = -\frac{\partial}{\partial \tilde{x}_k^0} \int_S \tilde{u} \, \beta_{ij} n_j \, dS. \tag{6a}$$

Writing out the integrand explicitly, and noting that $r = a$ over S, we have

$$I_2 = -\frac{\partial}{\partial \tilde{x}_k^0}\left(\tfrac{1}{2} U^2 a^3 \int_S \frac{n_i}{\tilde{r}} dS\right). \tag{7}$$

On making use of (1), (7) becomes

$$I_2 = \frac{2\pi}{3} U^2 a^6 \frac{\partial^2}{\partial \xi_i \partial \xi_k}\left(\frac{1}{\xi}\right). \tag{8}$$

Substituting (5) and (8) in (4.11), we have the desired result (4.12).

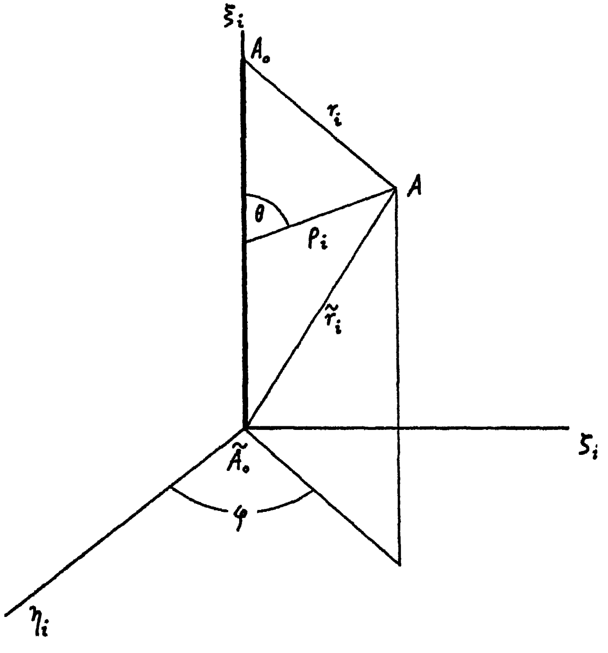


FIGURE 6.—Scheme of coordinate systems used for calculation—spherical coordinates (ρ, θ, φ) and bipolar coordinates $(r, \tilde{r}, \varphi)$.

Case (*2*). To evaluate the integrals (4.17) – (4.21), we use bipolar coordinates $(r, \tilde{r}, \varphi)$ (Fig. 6). In that figure, the spherical coordinates (ρ, θ, φ) are also shown. The quantities ξ_i and ρ_i are components of $\tilde{A}_0 A_0$ and OA referred to the general coordinate axes (not shown). We shall find the following formulae useful:

(i) The element of volume is given by

$$d\tau = \frac{r \tilde{r}\, dr\, d\tilde{r}\, d\varphi}{\xi}. \tag{9}$$

(ii) The element of area on the plane of symmetry $r = \bar{r}$ is

$$dS = \rho\, d\rho\, d\varphi = r\, dr\, d\varphi = \bar{r}\, d\bar{r}\, d\varphi. \tag{10}$$

(iii) The average values of ρ_i and $\rho_i\rho_k$ with respect to φ are given by

$$\frac{1}{2\pi}\int_0^{2\pi} \rho_i\, d\varphi = \rho \cos\theta \frac{\xi_i}{\xi}, \tag{11}$$

$$\frac{1}{2\pi}\int_0^{2\pi} \rho_i\rho_k\, d\varphi = \tfrac{1}{2}\rho^2 \left\{ \sin^2\theta\, \delta_{ik} + (2\cos^2\theta - \sin^2\theta)\frac{\xi_i\xi_k}{\xi^2} \right\}. \tag{11a}$$

Over the plane of symmetry, we have $\theta = \pi/2$, and hence

$$\frac{1}{2\pi}\int_0 \rho_i\, d\varphi = 0, \tag{12}$$

$$\frac{1}{2\pi}\int_0^{2\pi} \rho_i\rho_k\, d\phi = \frac{\rho^2}{2}\left(\delta_{ik} - \frac{\xi_i\xi_k}{\xi^2}\right). \tag{12a}$$

We have also for $\theta = \pi/2$

$$\frac{1}{2\pi}\int_0^{2\pi} \rho_i\rho_j\rho_k\, d\varphi = 0. \tag{12b}$$

To verify (11) and (11a), we write

$$\frac{\rho_i}{\rho} = \cos\theta\, \xi_i^{(0)} + \sin\theta \cos\varphi\, \eta_i^{(0)} + \sin\theta \sin\varphi\, \zeta_i^{(0)}, \tag{13}$$

where $\xi_i^{(0)}$, $\eta_i^{(0)}$, $\zeta_i^{(0)}$ are unit vectors along the directions of the vectors ξ_i, η_i, ζ_i (Fig. 6). When we substitute this expression in the integrands of (11), (11a), and make use of the relation

$$\xi_i^{(0)}\xi_k^{(0)} + \eta_i^{(0)}\eta_k^{(0)} + \zeta_i^{(0)}\zeta_k^{(0)} = \delta_{ik}, \tag{14}$$

the required results follow immediately, since $\xi_i^{(0)} = \xi_i/\xi$.

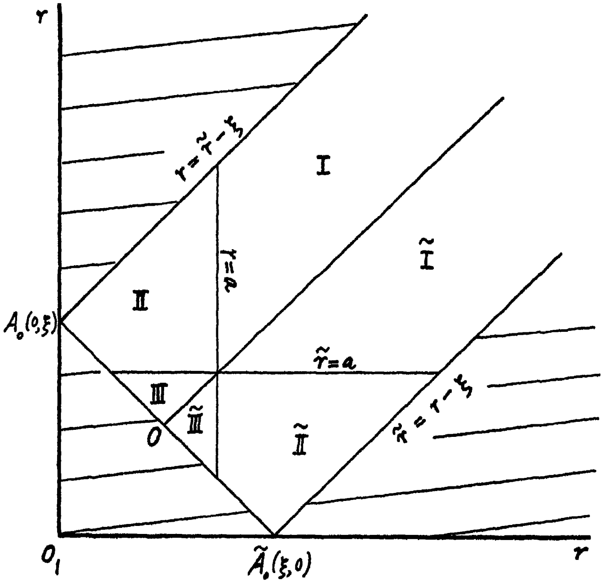
Other formulae to be used are

$$r_i = \rho_i - \frac{\xi_i}{2}, \quad \bar{r}_k = \rho_k + \frac{\xi_k}{2}, \tag{15}$$

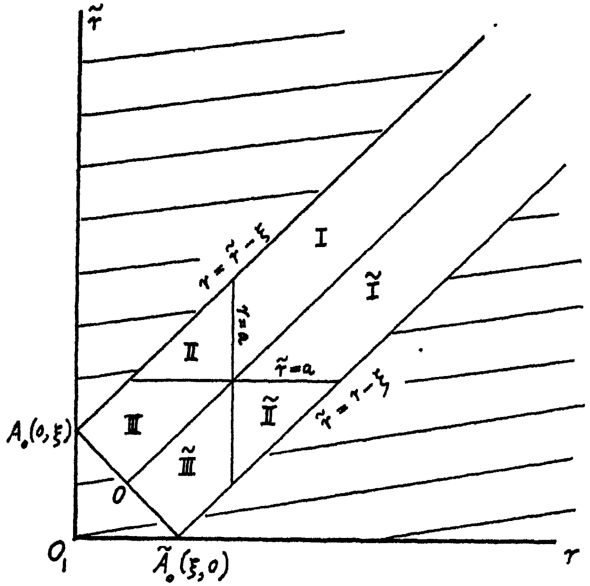
and

$$n_i = -\frac{\xi_i}{\xi}, \quad \rho^2 = r^2 - \left(\frac{\xi}{2}\right)^2, \quad \text{over the plane of symmetry.} \tag{16}$$

Equations (10), (12), (12a), (12b), (15) and (16) are the necessary formulae for evaluating the integrals (4.17), (4.18), (4.19) over the



Case (2a) $a \leq \xi \leq 2a$



Case (2b) $0 \leq \xi \leq a$

Figure 7.—Partition of space as appearing in the $(r, \tilde{r})$ plane (cf. Fig. 3; the shaded regions have no correspondence in space).

plane of symmetry. The computation is straightforward but a little lengthy. We need only note the limits of integration:

(i) φ is to be integrated over the interval $(0, 2\pi)$;

(ii) r is to be integrated over the range (a, ∞) for II′ and over $\left(\frac{\xi}{2}, a\right)$ for II″.

We may remark that the calculations in Case (1) can also be done by this scheme with r integrated over the range $\left(\frac{\xi}{2}, \infty\right)$ to cover the whole plane of symmetry.

The evaluation of the volume integral (4.21) is also straightforward. Only the limits of integration will be explained.

We first integrate with respect to φ from 0 to 2π, leaving a double integral over a region in the $(r, \tilde{r})$ plane. In Fig. 7, we show the various regions of that plane corresponding to the different portions of space $(I, II, III, \tilde{I}, \tilde{II}, \tilde{III})$ for both the Cases (2*a*) and (2*b*). The integral is therefore of the form:

$$I = \int_{\xi/2}^{a} d\tilde{r} \int_{\xi-\tilde{r}}^{\tilde{r}} dr\, f(r,\tilde{r}) \tag{17}$$

for Case (2*a*) and

$$I = \int_{\xi/2}^{\xi} d\tilde{r} \int_{\xi-\tilde{r}}^{\tilde{r}} dr\, f(r,\tilde{r}) + \int_{\xi}^{a} d\tilde{r} \int_{\tilde{r}-\xi}^{\tilde{r}} dr\, f(r,\tilde{r}) \tag{18}$$

for Case (2*b*). The integrand is found to be odd in r, and hence (17) and (18) are identical, the integral

$$\int_{\xi-\tilde{r}}^{\tilde{r}-\xi} f(r,\tilde{r})\, dr$$

being zero. The actual evaluation of (17) involves only the evaluation of the integrand (which can be easily done with the help of the formulae in §3) and the integration of certain polynomials.

REFERENCES

(1) Taylor, G. I. Proc. Lond. Math. Soc., 20: 196-212. 1921.

(2) Taylor, G. I. Proc. Roy. Soc., A 151: 421-478. 1935.

(3) Kármán, Th. von. J. Aero. Sci. 4: 131-138. 1937. Kármán, Th. von, and Howarth, L. Proc. Roy. Soc., A 164: 192-215. 1938.

(4) ROBERTSON, H. P. Proc. Camb. Phil. Soc., **36**: 209-223. 1940.
(5) BURGERS, J. M. K. Ned. Akad. Wet., Proc., **43**: 2-12. 1940.
(6) TAYLOR, G. I., and GREEN, A. E. Proc. Roy. Soc., A **158**: 499-521. 1937.
(7) HILL, M. J. M. Phil. Trans., A **185**: 213-245. 1894. See also Lamb, Hydrodynamics, p. 245, 1932; Villat, H., Leçons sur la Theorie des Tourbillons, pp. 195-204, 1930.
(8) TAYLOR, G. I. J. Aero. Sci., **4**: 311-315. 1937.
(9) MACPHAIL, D. C. J. Aero. Sci., **8**: 73-75. 1940.
(10) DRYDEN, H. L. Th. von Kármán Anniversary Volume, pp. 85-102. 1941.
(11) TAYLOR, G. I. Proc. Roy. Soc., A **164**: 476-490. 1938.

TRANSACTIONS OF

THE ROYAL SOCIETY OF CANADA

SECTION III

CHEMICAL, MATHEMATICAL, AND PHYSICAL SCIENCES

THIRD SERIES—VOLUME XXXVIII—SECTION III

MAY, 1944

OTTAWA
THE ROYAL SOCIETY OF CANADA
1944

TABLE OF CONTENTS

Reactions in the Solid State. By T. THORVALDSON, F.R.S.C. 1

Theory of Algebraic Functions based on the Use of Cycles. By S. BEATTY, F.R.S.C., and MURIEL WALES.............. 11

Transactions of the Royal Society of Canada
SECTION III

SERIES III MAY, 1944 VOL. XXXVIII

PRESIDENTIAL ADDRESS

REACTIONS IN THE SOLID STATE

By T. THORVALDSON, F.R.S.C.

INTRODUCTION

CHEMICAL reactions in gases and solutions were the first to be studied theoretically on account of their simple mechanisms. The atoms, molecules, or ions involved are relatively free and mobile and reactions between them lend themselves readily to quantitative treatment. The study of reactions involving solids introduces a number of new factors of considerable complexity and this makes quantitative treatment difficult. Industrially such reactions are of great importance and empirical knowledge has therefore generally been far in advance of theory.

The modern advances in the theory of reactions in the solid state were made possible by increased knowledge of crystal structure brought about by the use of *x*-rays. Starting with this firm foundation, the application of methods familiar to physical chemists and physicists for the measurement of rates of diffusion, electrical and thermal conductivity, photoelectric and photochemical effects, heats of reaction, absorption spectra, energies of activation, etc , has resulted in considerable theoretical progress during the last fifteen to twenty years. The use of the electron microscope in the study of such processes promises to lead to further advances.

The available quantitative results apply mainly to reactions between crystalline solids, and especially to reactions between crystals possessing ionic lattices. The study of cellulose and rubber, where the unit groups are constantly repeated, might in some ways be considered as a borderline field. Some properties of these are analogous to those of well-defined crystal lattices such as the lattices of graphite, solid solutions, zeolites, etc. Many of the conclusions as to reactions between crystalline substances are equally applicable to reactions in amorphous or glassy systems.

17091

DIFFUSION PHENOMENA

The classical determination of the rate of diffusion of gold into lead published by Roberts-Austen in 1896 may perhaps be considered as the first fundamental measurement of the kind applicable to the study of reaction rates in solids. The later studies of Tammann and others gave additional information as to the mobilities of particles in metallic systems and the formation of intermetallic compounds.

The simplest case of diffusion is the so-called "self-diffusion" or the penetration of a particle by diffusion into a system composed of chemically identical particles. This conception, which was introduced by Maxwell in a discussion of the kinetic theory of gases, is of great importance in connection with reactions in the solid state. Actual measurements of rates of self-diffusion could be made only after the discovery of radioactive isotopes and became of general application with the introduction of artificially radioactive isotopes as indicators.

Von Hevesy and his co-workers compared the rate of diffusion of gold into lead at 165° with the rate of self-diffusion of lead into lead using a radioactive lead isotope. They found that gold diffused five million times as fast as the lead. The ratio of diffusion rates was even greater at lower temperatures. Von Hevesy explained this difference in diffusion rates on the basis that the high symmetry in the distribution of electric charges in the pure metal gives a very stable system so that even near the melting point very few lead atoms have the kinetic energy necessary to free themselves from their neighbours. In the solid solution of gold in lead the binding of the gold atoms is much weaker and these are therefore more mobile at a given temperature. He supported his conclusion by pointing out that as gold is progressively substituted by metals chemically more and more similar to lead, the rate of diffusion in lead is progressively reduced. Others consider that the diameter of the atom or ion may be the main determining factor.

The change in the value of the coefficient of self-diffusion with temperature may be represented by an expression of the familiar form, $D = Ce^{-E/kT}$ where D is the diffusion coefficient, C a constant, E the activation energy of the process of diffusion and k the Boltzman constant. The value for E in the self-diffusion of lead is 28,050 cal. ("heat of loosening of the lattice particle"). Calculated values for the diffusion coefficient, D, agree well with experimental values.

By using artificially radioactive ions one can determine diffusion rates for the self-diffusion of positive and negative ions in their crystalline salts. From these data ionic mobilities and electrical conductivities of ionic crystals can be calculated by the method of Nernst. Diffusion rates can also be calculated from conductivity data for crystals.

Diffusion in anisotropic crystals presents some points of interest in relation to chemical reaction rates. Rates of diffusion for different crystal faces differ. Thus, rates of self-diffusion of bismuth atoms in hexagonal bismuth crystals parallel and perpendicular to the main axis are very different. Similar results have been obtained with diffusion of iodine ions in lead iodide, while the rate of diffusion of lead ions in lead iodide differs only slightly with direction. Hedvall has found differences in rates of chemical reaction on different faces of anisotropic crystals.

Electrical conductivity of solids is in general closely related to rate of diffusion and chemical reactivity. Conductivity of crystals may be entirely ionic or electronic or a mixture of the two. Electronic conductivity appears in many compounds such as some oxides, sulphides, selenides, nitrides, and carbides of the heavy metals. Ionic lattices in general show ionic conductivity, but many transition types are brought about by polarization effects. Some ionic crystals such as the silver halides and alkali halides at low temperatures are pure cationic conductors; others such as the chlorides and bromides of barium and lead are pure anionic conductors, while in some cases both ions are transported through the crystal as in alkali halides at high temperatures and in lead iodide. The type of conductance may vary gradually with temperature and in a discontinuous manner at transition points.

IRREGULARITIES IN THE STRUCTURE OF CRYSTALS AND ELECTRICAL CONDUCTIVITY

Neither the classical theory nor the early *x*-ray investigations of crystal structure considered the possibilities of imperfections in a normal crystal lattice. Later, when information as to lattice energies was available, theoretical calculations of the physical characteristics of ideal crystals, such as compressive strength and hardness, were made but the values obtained did not agree with those actually found. It appeared that real crystals possessed lattices much less rigid than would be expected for geometrically perfect crystals. Explanations suggested included: weakness due to presence of traces of foreign impurities, contractions in the lattice as a thermodynamic property of crystals, a dissociation equilibrium between charged and uncharged particles in the lattice, etc. Studies were made of the distribution of charges in lattices, types of bonds between lattice particles, polarization effects, ionic volumes and the relation of these to the symmetry and other properties of crystals.

The most generally accepted theory of electrical conductivity in ionic crystals is connected with the names of Frenkel, Wagner, Schottky, and Jost. Recurring imperfections are present in the lattices of crystals, the frequency of such disorder in any given structure in thermodynamic equilibrium being determined by the temperature. The thermal motion of the lattice particles causes some of them to leave their normal position and to be trapped in unsymmetrical positions in the crystal lattice. Thus, in an ionic crystal of a pure substance in thermal equilibrium at a given temperature, there will be a certain number of lattice points from which the ion is missing and except in special cases the same number of ions located between lattice points. The displaced or interstitial ions and vacant spaces increase in number with rising temperature and are mobile in an electric field. At low temperatures they become immobile. This may mean at room temperature for one substance (as sodium chloride) and at liquid-air temperature for another substance (as silver bromide).

Electrical conductivity in many solid ionic compounds is explained by this theory as being due to the movement of interstitial ions and vacant lattice points. Electrolytic conductivity may be due to the following:

(1) An ion jumps from one interstitial position to another.

(2) An interstitial ion takes the place of an ion in the lattice, the latter moving on to an interstitial position.

(3) An ion jumps from its lattice place to an adjacent vacant lattice point.

At temperatures of from 100° to 200° below the melting point there is in general a thermodynamic equilibrium of the interstitial ions and vacant places and the crystal therefore possesses definite conductivity. At lower temperature the conductivity of the solid depends on the thermal and mechanical history of the specimen.

The effect on electrical conductivity produced by increasing the number of vacant lattice points can be studied by means of solid solutions. Thus, if one forms a solid solution of cadmium chloride ($CdCl_2$) in silver chloride ($AgCl$), it is evident that one vacant positive lattice point will be produced for every molecule of cadmium chloride introduced into the lattice, since the ratio of the number of cations to anions is 1 to 2 in $CdCl_2$ and 2 to 2 in $AgCl$. It is found that the conductivity of the solid solution increases essentially in the same ratio as the number of vacant lattice points indicating that the conductivity is due mainly to the movement of cations through the vacant positions, or stated differently, due to the mobility of the vacant positions. The conductivity due to the interstitial silver ions *per se* is therefore negligible in this case. This gives a method for calculating the concentration of the number of vacant positions (and interstitial ions) in pure silver chloride at the same temperature. The chlorine ions apparently form a rigid lattice while the current is carried only by the cation, which shows disorder at 300° C. to the extent of about five-hundredths of one per cent. That means that one ion in two thousand is present in interstitial positions with the same proportion of vacant lattice points. The corresponding value for silver bromide is four per thousand.

In some cases such as the α-modifications of AgI and Ag_2S, x-ray investigations show a perfect anion lattice with the cations distributed in almost complete disorder. The cationic conductivity is extremely high, nearly as high as in the molten state.

Only cases where there is disorder of cations have thus far been considered. In other crystals such as KCl both the cations and anions contribute to the conductivity. Here, according to the theory, there is a certain amount of disorder among both cations and anions. In cases of substances showing pure anionic conductivity there is disorder only among the anions, and the cations form a perfect lattice. The main determining factors are the relative ionic diameters, the interstitial space and the binding forces between the ions.

This theory may represent an over-simplification of the mechanism of conductivity of solids even of the ionic type. Mott considers that an electronic current might exist in a polar crystal such as sodium chloride. If an electron is removed from a chloride ion we have a chlorine atom and a "positive hole" which is mobile, since an electron from an adjacent chloride ion can "fill it up" and this process may then be repeated through the crystal. Thus by a series of transferences of electrons and no transference of chloride ions, the chlorine atom may reach the anode. This may be in part the mechanism of photoelectric currents and of electrical conductivity in the case of certain semi-conductors. In an analogous manner, an extra electron introduced into the crystal lattice may transform a sodium ion into a sodium atom and then by a series of jumps of the electron from ion to ion the sodium atom may reach the cathode through the movement of an electron without actual transference of ions.

Further, given a lattice point from which a negative ion is missing an electron may be trapped in the potential "hole" giving rise to so-called "F centres" and "F spectrum bands" in the visible. An example is rock salt irradiated by x-rays, producing yellow coloration. Annealing of such crystals produces colloidal particles of sodium in the crystals. There is also the theoretical possibility of the trapping of two electrons in the potential hole and the trapping of electrons on metallic atoms, which will not be considered here. All these would have an influence on chemical reactions in the solid state, especially photochemical or oxidation-reduction reactions.

The theory in general lends itself to treatment by the methods of statistical mechanics.

The above discussion refers only to diffusion inside of single crystals. It is evident that a theory of diffusion based on disorderly arrangement in crystal lattices may have important applications in explaining physical and chemical phenomena occurring at surfaces and interfaces.

APPLICATIONS TO CHEMICAL REACTIONS BETWEEN SOLIDS

Wagner has applied these ideas to chemical reactions in the solid state with considerable success.

In a reaction, involving only solids, of the type: $A + B \rightarrow AB$, the reaction product, AB, is formed between the initial phases A and B. For further reaction diffusion of the reactants in the solid reaction product AB must take place. The rate of diffusion of A and B in AB is therefore the main factor determining the rate of reaction. Wagner has suggested that the diffusion represents ionic transport under a potential gradient rather than mere thermal diffusion. For some systems such as the system, $2AgI + HgI_2 \rightarrow Ag_2HgI_4$, this seems to be in agreement with experiment. The anion lattice remains fixed while both the silver and mercury ions migrate. The α-Ag_2HgI_4 phase can dissolve both silver iodide and mercuric iodide. In the product about every fourth positive lattice point is unoccupied. It is interesting to note that α-Ag_2HgI_4 possesses about 1,000 times the conductivity of any other solid substance studied in the temperature range 50° to 150°.

Wagner has also applied his theory to reactions of the type, $A + B \rightarrow C + D$, especially the reaction, $Cu + AgCl \rightarrow Ag + CuCl$. He has found that the reacted layer formed between the original phases is a disorderly conglomerate of Ag and CuCl. The chemical change can be calculated approximately from the e.m.f. of the galvanic cell, Ag/AgCl/CuCl/Cu (0.5 volts) considering transport numbers of the ions and the free energy decrease of the reaction. The method can be extended to double decompositions involving two salts.

Wagner's theory has been applied to the tarnishing of metals.

Such a system is represented as two single electrodes with the film of reaction product acting as both electrolyte and external circuit (it possesses both ionic and electronic conductances). The reaction is the reverse of electrolysis.

The low rate of tarnishing of pure aluminum and beryllium and certain of their alloys is explained on the basis of an oxide film of extremely high electrical resistance preventing ionic transport.

Among the earliest systematic studies of reactions between solids are those of Hedvall and his co-workers. They determined heating curves with reactive powder mixtures. Deviations in the temperature-time curve during constant rate of heating, from the curve obtained for a non-reacting powder of the same heat capacity and under the same conditions, indicates exothermic or endothermic reactions as the case may be.

The reactions studied were mainly exothermic displacement reactions of the type: $A'O + A''XO_n \rightarrow A''O + A'XO_n$. Powder mixtures of the three oxides, BaO, SrO and CaO with carbonates, sulphates, phosphates, and silicates of the alkaline earths and heavy metals were used. There was a very striking similarity in the temperature at the beginning of rapid reaction depending on the free oxide and almost independent of the salt used. Thus the temperatures of rapid reaction, with only minor deviations, were: for barium oxide (BaO) about 350°, for strontium oxide (SrO) about 450°, for calcium oxide (CaO) about 530°.

Hedvall and his co-workers found certain interesting exceptions to this rule. The three oxides when heated with silver nitrate all gave approximately the same temperature of rapid reaction, namely 170°, which is very much lower than the reaction temperature of any of the oxides with the other salts used. Crystalline silver nitrate has a transition point at 160° and this led Hedvall to study reaction rates with reference to transition temperatures of the reactants. Using crystalline silver sulphate which has a transition point at 411°, he found that the reaction temperature with BaO was 340°, which is approximately the same as with the other salts and is well below the transition point of 411°, but both SrO and CaO reacted rapidly at 420° which is just above the transition point of silver sulphate but well below the normal reaction temperature for these oxides.

This high reactivity of solids at transition points (or rather approximately at the temperature at which the rate of transition is at a maximum) appears to be general, the reaction rate being lower both below and above this temperature. Jost, however, states that in some cases there is no maximum in the curve for the rate of reaction but that there is an increase in reactivity at the transition temperature because the high temperature phase has a higher reactivity. There is some experimental evidence that solid solutions of metals and alloys show similar maxima for reaction rates at transition points.

Hedvall comes to the conclusion that due to the disturbed condition of the crystal lattice at the transition point, the activation energy of the diffusion process ("energy of loosening") is very small and the reactivity therefore high. He considers that in his displacement reactions with the oxides of barium, strontium, and calcium and salts of oxy-acids, one is dealing with diffusion of the oxide and not of metallic ions and that the predominating factor (in the absence of a low transition point for the salt) is therefore the rate of diffusion of the oxide used.

On the other hand, in Hedvall's experiments with barium oxide and the halides of copper, the reaction temperature was not the same for the different halides, but agreed with the temperature at which Tubandt found a sudden increase in the ionic conductivity of the respective halides. Here the reaction is apparently due to ionic migration.

Jander and his co-workers carried out many experiments using compact pellets of crystalline powders, determining material transport by analysis. In the reaction, $ZnO.WO_3 + MgO \rightarrow ZnO + MgO.WO_3$, they found that only the tungstic oxide diffused across the boundary. Many other similar cases were found. Pairs of substances behaving in this way generally showed pure electronic conductivity. In some cases, however, they found mixed conductivity and in others they concluded that the reactions took place through diffusion of ions rather than anhydride groups.

Tammann, Jander, and others have proposed mathematical expressions for rates of reaction between solid phases. Lack of time prevents consideration of these. The same applies to any discussion of ultrasonic, magnetic, photoelectric, or photochemical effects on reactions in the solid state. This brief account is confined mainly to reactions of normal or thermodynamically stable phases. Consideration of effects on rates of reaction of surface phenomena and phase boundaries, imperfect lattices and metastable systems produced by mechanical and other means is also omitted. These are of extreme importance in connection with metallic systems, recrystallization and annealing of metals, age hardening of alloys, sintering processes, and heterogeneous catalysis. Reactions of solid explosives and detonators is another field which is not considered.

In conclusion, the reaction between calcium oxide and silica, which on account of its technical importance has been given a rather detailed study, especially by Jander, will be considered briefly. Four different products are possible, namely compounds with molar ratios of lime to silica of 1:1, 3:2, 2:1, and 3:1. When calcium oxide and quartz are brought together in any of these proportions and heated at 1000° to 1200°, it appears that the compound with the ratio 2:1 is at first formed. The proportion of lime and silica thus does not appear to affect the initial reaction, although it does determine the product obtained on prolonged heating. As the melting points of calcium oxide and of the different modifications of silica are well above this temperature and the lowest eutectic in the binary system is at 1436°, there can be no question of a liquid phase at 1000 to 1200° C.

A description of what happens when a 1 to 1 mixture of lime and silica is heated at 1200° C, gives a good picture of a rather complicated reaction. The first product which can be detected in appreciable quantity is dicalcium silicate. This does not necessarily mean that this product is formed in a single step. It is possible that monocalcium silicate is first formed but that due to the chemical activity of the nascent molecule of calcium silicate and the greater mobility of calcium oxide than of silica, a second molecule of lime is immediately taken up to form dicalcium silicate. The high rate of diffusion of calcium oxide gives an excess of this substance in the film of product formed between the two phases and causes the reaction to take place on the silica side of the film. As the reacted film becomes thicker, the supply of lime to the silica side of the film is slowed down and silica molecules in the film become relatively more plentiful. At this stage the x-ray patterns begin to indicate presence of the 3:2 calcium silicate in the mixture, the proportion increasing with time. With further decrease in the supply of lime, the x-ray patterns begin to show the presence of monocalcium silicate and the dicalcium silicate begins to disappear. Later the dicalcium silicate has completely disappeared and the proportion of the monocalcium silicate continues to increase at the expense of the 3:2 compound until after prolonged heating the pattern shows only the monocalcium silicate. None of the 3:1 compound, tricalcium silicate, is formed as this is unstable at 1200° C.

When the ratio of CaO to SiO_2 in the original reaction mixture is increased, the initial observations are the same, mixtures of silicates being formed with reversal to the compounds nearest the original composition of the mixture. If the ratio is between 2 to 1 and 3 to 1 and the temperature is in the range 1250° to 1900° where tricalcium silicate is stable, the dicalcium silicate and the excess of lime react to form tricalcium silicate until all the free lime has disappeared.

Thus it is evident that the relative rates of diffusion of the reacting units and the stability of the possible compounds at the temperature used determine the progress of the reaction. The picture is, however, only an "over-all" one and does not settle definitelv whether ions or molecules are the diffusing units. It is unlikely that in the formation of the silicates the silicon and calcium ions are the diffusing units, but the reaction could also be explained on that basis. In similar reactions between MgO and Al_2O_3 to form spinel, it is probable that the diffusing units are the small cations of magnesium and aluminum.

In conclusion one might say that while a good beginning has been made towards explaining the mechanism of reactions in the solid state, apparently no satisfactory general theory has so far appeared. Diffusion is definitely the most important factor in determining both the rate and mechanism of such reactions. The theory that diffusion in the solid state is essentially a movement of ions under a potential gradient is the most interesting and promising one, but its application is doubtful in many cases where it appears likely that the diffusing particles are neutral molecular groups.

THEORY OF ALGEBRAIC FUNCTIONS BASED ON THE USE OF CYCLES

By S. BEATTY, F.R.S.C., and MURIEL WALES

1. *Introduction*

Perhaps no topic in Mathematics exhibits as wide a variety of treatment as the theory of algebraic functions associated with a given fundamental equation $F(z, u) = 0$. Indeed, among purely algebraic treatments, as distinct from those that are geometric or transcendental, there is one that rests on a study of the form of algebraic functions relative to all fine points $z = a$, taken at once, and another that is concerned with the form relative to each finite point $z = a$, taken by itself. All existing treatments make use of the branches of u, appearing as series in powers of the element $z - a$ for the finite point $z = a$ and the element $\frac{1}{z}$ for the point at infinity. In the present paper, it is proposed to develop the theory by making use of the cycles rather than the branches at each point. Fractional powers of the element are present in a branch, except when the branch is also a cycle, but, of course, only integral powers of the element appear in a cycle. We shall outline a process for obtaining the cycles at a point. Where a cycle is made up of several branches, it will, in general, be supplied by the process to be described, and in that case it will not be necessary to extend it to supply the constituent branches. It may happen, however, that the process does not supply the cycles directly, in which case it must be extended to furnish the branches, which can then be combined to yield the cycles. The method for obtaining the cycle factorization of $F(z, u)$ relative to a given point was suggested by that employed in an earlier paper for obtaining the branch factorization(1). Here, as before, the given point will be $z = 0$.

2. *The generalized resultant*

Let $P_i(x) \equiv x^{n_i} + a_i^{(1)} x^{n_i-1} + \ldots. + a_i^{(n_i)}$, $(i = 1, 2, \ldots., m)$, denote m polynomials in x, of degrees $n_1, n_2, \ldots., n_m$, and let $P(x)$ denote their product, of degree $N = n_1 + n_2 + \ldots + n_m$. Construct the m polynomials

$$Q_i(x) \equiv \frac{P(x)}{P_i(x)}, \ (i = 1, 2, \ldots., m),$$

and form the N polynomials, in order,

$$x^{n_i-1}Q_i(x),\ x^{n_i-2}Q_i(x), \ldots.,\ Q_i(x),\ (i = 1, 2, \ldots., m).$$

The determinant of the coefficients of these N polynomials, with rows referring to polynomials in the given order and columns to powers of x in descending order, will be called the generalized resultant of the polynomials $P_1(x), P_2(x), \ldots\ldots, P_m(x)$ and denoted by

$$D = R(P_1, P_2, \ldots\ldots, P_m).$$

We call attention to the following properties of $R(P_1, P_2, \ldots\ldots, P_m)$:

(i) $R(P_1, P_2, \ldots\ldots, P_m)$ is equal numerically to the product of all the ordinary resultants $R(P_i, P_j)$, for $i < j$.

(ii) $R(P_1, P_2, \ldots\ldots, P_m)$ is zero when and only when some pair $P_i(x)$, $P_j(x)$ have a common root.

(iii) Where $\bar{P}(x)$ is a polynomial in x, of degree $N - 1$, there exist unique polynomials in x

$$\bar{P}_i(x), \text{ of degree } n_i - 1, \quad (i = 1, 2, \ldots\ldots, m),$$

such that

$$\bar{P}(x) \equiv \bar{P}_1(x)Q_1(x) + \bar{P}_2(x)Q_2(x) + \ldots\ldots + \bar{P}_m(x)Q_m(x),$$

provided that the generalized resultant $R(P_1, P_2, \ldots\ldots, P_m)$ is different from zero.

Property (ii) follows from property (i), on making use of the fact that each $R(P_i, P_j)$ is different from zero except when $P_i(x)$, $P_j(x)$ have a common root. Property (iii) is seen to hold, since $R(P_1, P_2, \ldots\ldots, P_m)$ is the fundamental determinant involved in solving for the coefficients in $\bar{P}_1(x), \bar{P}_2(x), \ldots\ldots, \bar{P}_m(x)$ so as to make the sum of their products by $Q_1(x), Q_2(x), \ldots\ldots, Q_m(x)$ respectively agree identically with $\bar{P}(x)$, and since this fundamental determinant is by hypothesis different from zero. It only remains, therefore, to verify that property (i) holds. Let x_i be a typical one of the n_i roots of $P_i(x)$, for $i = 1, 2, \ldots\ldots, m$. The coefficients of each $P_i(x)$ are integral, rational, symmetric functions of the n_i roots (x_i), and so $R(P_1, P_2, \ldots\ldots, P_m)$ is an integral, rational function of the N roots $(x_1), (x_2), \ldots\ldots, (x_m)$. Since the generalized resultant $R(P_1, P_2, \ldots\ldots, P_m)$ is zero if $x_i = x_j$, it must contain as a factor the product expression

$$\underset{i<j}{\pi}\ (x_i - x_j),$$

which is homogeneous in the N roots $(x_1), (x_2), \ldots\ldots, (x_m)$ and of degree

$$\sum_{i<j} n_i\, n_j$$

in them. But, if the rows of the generalized resultant $R(P_1, P_2, \ldots\ldots, P_m)$ are multiplied in order by

$$1, t, \ldots\ldots, t^{n_1-1}, 1, t, \ldots\ldots, t^{n_2-1}, \ldots\ldots\ldots, 1, t, \ldots\ldots, t^{n_m-1},$$

the result is a determinant Δ, in which the N columns in order are made up of terms of degrees $0, 1, \ldots, N-1$ in t, $(x_1), (x_2), \ldots, (x_m)$. It follows that Δ itself, in its expanded form, is made up of terms of degree $\frac{N(N-1)}{2}$ in t, $(x_1), (x_2), \ldots, (x_m)$. But, $\Delta = t^{\delta} R(P_1, P_2, \ldots, P_m)$, where

$$\delta = \sum \frac{n_i(n_i - 1)}{2},$$

and so $R(P_1, P_2, \ldots, P_m)$, in its expanded form, is made up of terms of degree

$$\frac{N(N-1)}{2} - \sum \frac{n_i(n_i - 1)}{2} = \sum_{i<j} n_i \, n_j$$

in $(x_1), (x_2), \ldots, (x_m)$. It follows, therefore, that

$$R(P_1, P_2, \ldots, P_m) = c \underset{i<j}{\pi} (x_i - x_j),$$

in which c denotes a constant depending for its value on the sequence $n_1, n_2, \ldots, n_m$ but independent of the particular roots $(x_1), (x_2), \ldots, (x_m)$ going with this sequence. We shall now determine the value of c. We may write, taking $P_1(x)$ to be x^{n_1},

$$R(x^{n_1}, P_2, P_3, \ldots, P_m) = c \underset{1<k}{\pi} (0 - x_k) \underset{1<i<j}{\pi} (x_i - x_j),$$

where the portion $\pi(0 - x_k)$, for a fixed k, involves $n_1 n_k$ factors. Writing $R(P_2, P_3, \ldots, P_m)$ in the form

$$\bar{c} \underset{1<i<j}{\pi} (x_i - x_j),$$

and applying Laplace's rule to write $R(x^{n_1}, P_2, P_3, \ldots, P_m)$ in the form

$$(-1)^{n_1(n_2 + n_3 + \ldots + n_m)} \left\{\pi(-x_k)\right\}^{n_1} R(P_2, P_3, \ldots, P_m),$$

we deduce by equating equivalent expressions for $R(x^{n_1}, P_2, P_3, \ldots, P_m)$ that

$$c = \bar{c}\,(-1)^{n_1(n_2 + n_3 + \ldots + n_m)}.$$

Using results of this type, as well as the particular result

$$R(P_{m-1}, P_m) = (-1)^{n_{m-1} n_m} \pi (x_{m-1} - x_m)$$

for the ordinary resultant, we find that

$$c = (-1)^q,$$

where q can be written in either of the forms

$$(n_1)n_2 + (n_1 + n_2)n_3 + \ldots\ldots\ldots + (n_1 + n_2 + \ldots + n_{m-1})n_m,$$

or

$$n_1(n_2 + n_3 + \ldots + n_m) + n_2(n_3 + \ldots + n_m) + \ldots\ldots + n_{m-1}(n_m),$$

which completes the verification of property (i).

3. *Cycle factorization of* $F(z, u)$

Let the fundamental equation $F(z, u) = 0$ have the form

$$F \equiv u^N + u^{N-1} f_1 + u^{N-2} f_2 + \ldots\ldots + f_N = 0,$$

in which $f_1, f_2, \ldots\ldots, f_N$ are rational functions of z. Denote by λ the least of the orders at the point $z = 0$ of any of the radicals

$$f_1^{\frac{1}{1}},\ f_2^{\frac{1}{2}}, \ldots\ldots, f_N^{\frac{1}{N}}.$$

We, therefore, have

$$F \equiv u^N + u^{N-1}(a_1 z^{\lambda} + \ldots) + u^{N-2}(a_2 z^{2\lambda} + \ldots) + \ldots + (a_N z^{N\lambda} + \ldots),$$

which involves only integral powers of z, and in which one at least of $a_1, a_2, \ldots\ldots, a_N$ is different from zero. Consider the expression

$$F_1 \equiv u^N + u^{N-1} a_1 z^{\lambda} + u^{N-2} a_2 z^{2\lambda} + \ldots\ldots + a_N z^{N\lambda},$$

made up of items each of weight $N\lambda$ in z, where u itself is regarded as having weight λ in z. In $F - F_1$, each item is of greater weight than $N\lambda$ in z. Let the least such weight be $N\lambda + \lambda_1 - \lambda$. Denote by G_1 the aggregate of items present in $F - F_1$ and having weight $N\lambda + \lambda_1 - \lambda$ in z. We may, therefore, write

$$G_1 \equiv (u^{N-1} b_1 z^{\lambda} + u^{N-2} b_2 z^{2\lambda} + \ldots\ldots + b_N z^{N\lambda})\, z^{\lambda_1 - \lambda},$$

in which one at least of $b_1, b_2, \ldots\ldots, b_N$ is different from zero, any particular b_i being certainly zero, however, if $i\lambda + \lambda_1 - \lambda$ is a fraction. We, therefore, have

$$F \equiv F_1 + G_1 + \text{items of greater weight than } N\lambda + \lambda_1 - \lambda \text{ in } z.$$

Let the different irreducible factors of F_1 be denoted by $I_1, I_2, \ldots\ldots, I_m$, each involving only integral powers of z. We may, therefore, write F_1 in the form

$$F_1 \equiv I_1^{p_1} I_2^{p_2} \ldots\ldots I_m^{p_m}.$$

We shall now investigate the possible forms of $I_1, I_2, \ldots\ldots, I_m$, depending on whether λ is integral or fractional. In case λ is an integer, it is clear that

$$I_1 \equiv u + C_1 z^{\lambda},\ I_2 \equiv u + C_2 z^{\lambda}, \ldots\ldots, I_m \equiv u + C_m z^{\lambda},$$

in which $C_1, C_2, \ldots\ldots, C_m$ are all different constants, of which one may be zero and one at least is different from zero. In case λ is a fraction $\frac{l}{n}$ in its lowest terms, it may happen that one of the irreducible factors $I_1, I_2, \ldots\ldots, I_m$ does not involve z, which means that this particular irreducible factor is u itself, but any irreducible factor I_i actually involving z is of the form $u^n + C_i z^{n\lambda}$, in which the constant C_i is different from zero. This is evident, since the only possible form that I_i could have is

$$u^{hn} + u^{(h-1)n} A z^{n\lambda} + u^{(h-2)n} B z^{2n\lambda} + \ldots . + H z^{hn\lambda},$$

which is reducible for h greater than unity. In case λ is a fraction $\frac{l}{n}$ in its lowest terms, the possibilities are, therefore, that either

$$I_1 \equiv u,\ I_2 \equiv u^n + C_2 z^{n\lambda}, \ldots . ,\ I_m \equiv u^n + C_m z^{n\lambda},$$

with the constants $C_2, \ldots . . ,\ C_m$ all different from zero and from one another, or

$$I_1 \equiv u^n + C_1 z^{n\lambda},\ I_2 \equiv u^n + C_2 z^{n\lambda}, \ldots . ,\ I_m \equiv u^n + C_m z^{n\lambda},$$

with the constants C_1, $C_2, \ldots . ,\ C_m$ all different from zero and from one another.

We recall that the first step in carrying out the cycle factorization of F was to write it in the form $F \equiv F_1 + G_1 +$ items of greater weight than $N\lambda + \lambda_1 - \lambda$ in z, where the form of the first factored part F_1, has been already noted. We shall now show the second step in the cycle factorization of F. We amplify the first factored part F_1 to obtain the second factored part F_2, namely

$$F_2 \equiv (I_1^{p_1} + J_{11} z^{\lambda_1 - \lambda})\ (I_2^{p_2} + J_{21} z^{\lambda_1 - \lambda}) \ldots . (I_m^{p_m} + J_{m1} z^{\lambda_1 - \lambda}),$$

where the functions

$$J_{11} z^{\lambda_1 - \lambda},\ J_{21} z^{\lambda_1 - \lambda}, \ldots . J_{m1} z^{\lambda_1 - \lambda}$$

are of less degrees in u than

$$I_1^{p_1},\ I_2^{p_2}, \ldots . ,\ I_m^{p_m}$$

respectively and are determined from the identity

$$G_1 \equiv F_1 (J_{11} z^{\lambda_1 - \lambda} I_1^{-p_1} + J_{21} z^{\lambda_1 - \lambda} I_2^{-p_2} + \ldots . + J_{m1} z^{\lambda_1 - \lambda} I_m^{-p_m}),$$

which determination is provided for by property (iii) relative to the generalized resultant $R(I_1^{p_1}, I_2^{p_2}, \ldots ., I_m^{p_m})$, taken with respect to u, which generalized resultant has a value of the form $Cz^{K\lambda}$, where the constant C is different from zero and $K\lambda$ is an integer. It appears that the second factored part F_2 involves the first factored part F_1, as well as G_1, and additional items of greater weight than $N\lambda + \lambda_1 - \lambda$ in z. It should be noted that some of the functions

$$J_{11} z^{\lambda_1 - \lambda},\ J_{21} z^{\lambda_1 - \lambda}, \ldots . ,\ J_{m1} z^{\lambda_1 - \lambda}$$

may be identically zero, but one at least must be different from zero. Just as G_1 was determined as the aggregate of items of least weight $N\lambda + \lambda_1 - \lambda$ in z present in $F - F_1$, let G_2, given by

$$G_2 \equiv (u^{N-1} c_1 z^{\lambda} + u^{N-2} c_2 z^{2\lambda} + \ldots . + c_N z^{N\lambda})\ z^{\lambda_2 - \lambda},$$

be determined as the aggregate of items of least weight $N\lambda + \lambda_2 - \lambda$ in z present in $F - F_2$. It is, of course, plain that λ_2 exceeds λ_1.

The result of the second step is, therefore, to write F in the form

$$F \equiv F_2 + G_2 + \text{items of greater weight than } N\lambda + \lambda_2 - \lambda \text{ in } z,$$

where the form of the second factored part F_2 has already been noted.

The third step towards effecting the cycle factorization of F carries on the process begun in the first step and appearing in fully developed form in the second. For this purpose, we amplify the second factored part F_2 to obtain the third factored part F_3, given by

$$F_3 \equiv (I_1^{p_1} + J_{11} z^{\lambda_1 - \lambda} + J_{12} z^{\lambda_2 - \lambda}) \ldots . (I_m^{p_m} + J_{m1} z^{\lambda_1 - \lambda} + J_{m2} z^{\lambda_2 - \lambda}),$$

where the functions

$$J_{12} z^{\lambda_2 - \lambda}, \ldots . , J_{m2} z^{\lambda_2 - \lambda}$$

are of less degress in u than

$$I_1^{p_1}, \ldots \ldots \ldots \ldots , I_m^{p_m}$$

respectively and are determined from the identity

$$G_2 \equiv F_1 (J_{12} z^{\lambda_2 - \lambda} I_1^{-p_1} + \ldots . + J_{m2} z^{\lambda_2 - \lambda} I_m^{-p_m}),$$

which determination is assured by property (iii) relative to the generalized resultant $R(I_1^{p_1}, \ldots . , I_m^{p_m}) = Cz^{K\lambda}$. It appears that the third factored part F_3 involves the second factored part F_2, as well as G_2, and additional items of greater weight than $N\lambda + \lambda_2 - \lambda$ in z. It is plain that some but not all of the functions

$$J_{12} z^{\lambda_2 - \lambda}, \ldots . , J_{m2} z^{\lambda_2 - \lambda}$$

may be identically zero. Let now G_3, given by

$$G_3 \equiv (u^{N-1} d_1 z^{\lambda} + u^{N-2} d_2 z^{2\lambda} + \ldots . + d_N z^{N\lambda}) z^{\lambda_3 - \lambda},$$

be determined as the aggregate of items of least weight $N\lambda + \lambda_3 - \lambda$ in z present in $F - F_3$, where, of course λ_3 exceeds λ_2. The result of the third step is, therefore, to write F in the form

$$F \equiv F_3 + G_3 + \text{items of greater weight than } N\lambda + \lambda_3 - \lambda \text{ in } z,$$

where the form of the third factored part F_3 has been already noted.

It is plain that the process continues indefinitely through successive steps and yields the factorization of F in the form

$$F \equiv \pi (I_i^{p_i} + J_{i1} z^{\lambda_1 - \lambda} + J_{i2} z^{\lambda_2 - \lambda} + \ldots . .),$$

with only integral powers of z appearing, even though some or all of $\lambda, \lambda_1, \lambda_2, \lambda_3, \ldots . .$ may be fractional. We shall refer to this as the process of formal factorization.

We must now take stock of the progress which all this implies towards obtaining the cycle factorization of F, and for this purpose we need consider only the factors of F already obtained, of which

$$H \equiv I^p + J_1 z^{\lambda_1 - \lambda} + J_2 z^{\lambda_2 - \lambda} + \ldots \ldots$$

is typical. The possibilities are as follows:

(i) λ is integral, with I having the form $u + Cz^{\lambda}$, in which the constant C is different from zero except possibly for a single one of the factors H.

(ii) λ is fractional, with I having the form u, which can be the case for at most one of the factors H.

(iii) λ is fractional, with I having the form $u^{n} + Cz^{n\lambda}$, in which the constant C is different from zero.

In (i), if $p = 1$, we see that $H = 0$ is already a cycle in z, of order 1, while, if $p > 1$, we make the substitution $u + Cz^{\lambda} = v$, which converts H into an expression of the form

$$H \equiv v^{p} + v^{p-1}h_1 + v^{p-2}h_2 + \ldots . + h_p,$$

where each h_i involves only integral powers of z, where the degree of H in v is likely to be less than the degree of F in u, and where the λ number of H with respect to v is certainly greater than the λ number of F with respect to u. The type of discussion already given for

$$F \equiv u^{N} + u^{N-1}f_1 + \ldots . + f_N$$

can then be carried out on

$$H \equiv v^{p} + v^{p-1}h_1 + \ldots . + h_p,$$

and this implies that all difficulties implied by possibility (i) can be met as they arise.

In (ii), the discussion deals with two sub-cases of the same type as those appearing under (i). In the second sub-case, the variable v is really u itself. This means, therefore, that all difficulties implied by possibility (ii) can be met as they arise.

In (iii), if $p = 1$, $H = 0$ is already a cycle in z, composed of the n branches

$$H \equiv (u - \epsilon c z^{\lambda} + \ldots)(u - \epsilon^{3} c z^{\lambda} + \ldots) \ldots . (u - \epsilon^{2n-1} c z^{\lambda} + \ldots),$$

where ϵ denotes $e^{\frac{\pi i}{n}}$ and c is written for $C^{\frac{1}{n}}$. However, if $p > 1$, it is necessary to make one or more substitutions and work down towards branches. We note that H is not just I^{p}, since this would mean that F too had repeated factors, which is a thing that would have been guarded against at the outset by using essentially the quotient of F by $R\left(F, \dfrac{\partial F}{\partial u}\right)$ in place of F itself. Writing ξ for $z^{\frac{1}{n}}$ and c for $C^{\frac{1}{n}}$, we see that the expression

$$I^{p} \equiv (u^{n} + Cz^{n\lambda})^{p}$$

converts into

$$I^{p} \equiv (u^{n} + c^{n}\xi^{n^{2}\lambda})^{p},$$

which is made up of all the items of least weight $n^{2}\lambda p$ in ξ present in H, where, of course, u must be regarded as having integral weight $n\lambda$ in ξ. In other words,

$$H_1 \equiv (u^n + c^n \xi^{n^2\lambda})^p$$

is the first factored part of H relative to ξ, just as F_1 was the first factored part of F relative to z. Even as F_1 was written as the product of powers of irreducible factors

$$F_1 \equiv I_1^{p_1} I_2^{p_2} \ldots . I_m^{p_m},$$

none involving fractional powers of z, so H_1 may be written as the product of powers of irreducible factors

$$H_1 \equiv (u - \epsilon c\, \xi^{n\lambda})^p\, (u - \epsilon^3 c\, \xi^{n\lambda})^p \ldots . (u - \epsilon^{2n-1} c\, \xi^{n\lambda})^p,$$

none involving fractional powers of ξ, and all being of the simplest form, namely linear in u. The process of formal factorization thus begun on H, expressed in terms of u, ξ, can be carried out to its conclusion so as to supply the corresponding factored form of H, namely

$$H \equiv \left\{(u - \epsilon c\, \xi^{n\lambda})^p + \ldots\right\} \left\{(u - \epsilon^3 c\, \xi^{n\lambda})^p + \ldots\right\} \ldots . \left\{(u - \epsilon^{2n-1} c\, \xi^{n\lambda})^p + \ldots\right\}.$$

Each of these n factors can be expressed in terms of v, ξ by making on them the substitutions

$$u - \epsilon c\, \xi^{n\lambda} = v,\ u - \epsilon^3 c\, \xi^{n\lambda} = v,\ \ldots . ,\ u - \epsilon^{2n-1} c\, \xi^{n\lambda} = v$$

in order, and then the process of formal factorization can be applied to them as functions of v, ξ.

On repeating the process of formal factorization, making use of substitutions for one or both variables as necessary, we observe that the cycles in z of $H = 0$ could be built up if the cycles in ξ of one of the n factors equated to zero were known, and in this way we could make use of induction to obtain the cycles in z of $H = 0$. For, if one of the n factors equated to zero is the product of q cycles in ξ, of orders $m_1, m_2, \ldots . , m_q$, then $H = 0$ is the product of q cycles in z, of orders $nm_1, nm_2, \ldots . , nm_q$, since a cycle in ξ of one of the n factors equated to zero combines with the $n - 1$ conjugate cycles in ξ of the remaining $n - 1$ factors equated to zero to constitute a cycle in z of $H = 0$. We can, therefore, count on obtaining the cycle factorization in z of $F = 0$.

4. *Function basis for a cycle equivalent to an assigned order*

Let $G \equiv u^\nu + u^{\nu-1} g_1 + \ldots . + g_\nu = 0$ denote a single cycle in z of $F = 0$. The corresponding branches are obtained by expressing G as a product of linear factors

$$G \equiv (u - u_1)\,(u - u_2) \ldots . (u - u_\nu),$$

whence the branches are seen to be $u = u_1, u = u_2, \ldots, u = u_\nu$, where $u_1 = \theta(\xi), u_2 = \theta(\epsilon\xi), \ldots., u_\nu = \theta(\epsilon^{\nu-1}\xi)$, in which ξ denotes $z^{\frac{1}{\nu}}$ and $\epsilon = e^{\frac{2\pi i}{\nu}}$, and $\theta(\xi)$ represents a power series in ξ.

In this connection, the fundamental theorem asserts the existence of $w = \rho(z, u)$, a rational function of (z, u) such that the ν expansions $w_1 = \rho(z, u_1), w_2 = \rho(z, u_2), \ldots., w_\nu = \rho(z, u_\nu)$ are each of order 1 in ξ, or, what is the same thing, the order of their product is 1 in z, or, otherwise expressed, $R(G, w)$ is of order 1 in z. The distinctive feature in the proof which we shall now give of the fundamental theorem is that simplicity is secured by confining the discussion to a single cycle and making use of determinants throughout.

The function w, defined by the Lagrange interpolation formula

$$D \equiv \begin{vmatrix} 1 & u & u^2 \ldots\ldots & u^{\nu-1} & w \\ 1 & u_1 & u_1^2 \ldots\ldots & u_1^{\nu-1} & \xi \\ 1 & u_2 & u_2^2 \ldots\ldots & u_2^{\nu-1} & \epsilon\xi \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & u_\nu & u_\nu^2 \ldots\ldots & u_\nu^{\nu-1} & \epsilon^{\nu-1}\xi \end{vmatrix} = 0,$$

is a polynomial of degree $\nu - 1$ in u, and the coefficients of $1, u, u^2, \ldots., u^{\nu-1}$ in this polynomial are power series in ξ. But, these power series in ξ are really power series in z, since if we suppose that one of them contains a power of ξ which is not a multiple of ν we must conclude that w changes its form if $\epsilon\xi$ replaces ξ throughout, which is opposed to the obvious fact that no change takes place in w as a result of $\epsilon\xi$ replacing ξ. Nevertheless, these power series in z need not be rational functions of z, and so w need not be of the form $\rho(z, u)$, a rational function of (z, u). Apart from this, the function w satisfies all of our requirements, since it takes the values $\xi, \epsilon\xi, \ldots., \epsilon^{\nu-1}\xi$ as u takes the values $u_1, u_2, \ldots., u_\nu$ and so has a total order of 1 in z for the cycle $G = 0$. This difficulty is removed, however, by replacing the power series $u_1, u_2, \ldots., u_\nu$, as they appear in D, by truncated power series $\bar{u}_1, \bar{u}_2, \ldots., \bar{u}_\nu$, obtained from them by deleting all powers of ξ after a sufficiently high power. It is plain that the ν values of w, namely $w_1, w_2, \ldots.., w_\nu$, still begin with the terms $\xi, \epsilon\xi, \ldots., \epsilon^{\nu-1}\xi$, and, of course, w is now of the required form $\rho(z, u)$, a rational function of (z, u).

The preceding discussion may be modified slightly to relate more explicitly to the cycle $G = 0$ and less to its branches. Let Δ denote the determinant

$$\Delta \equiv \begin{vmatrix} 1 & 0 & 0 & 0 & \ldots\ldots & 0 \\ 0 & 1 & 1 & 1 & \ldots\ldots & 1 \\ 0 & u_1 & u_2 & u_3 & \ldots\ldots & u_\nu \\ 0 & u_1^2 & u_2^2 & u_3^2 & \ldots\ldots & u_\nu^2 \\ \ldots & \ldots & \ldots & \ldots & \ldots\ldots & \ldots \\ 0 & u_1^{\nu-1} & u_2^{\nu-1} & u_3^{\nu-1} & \ldots\ldots & u_\nu^{\nu-1} \end{vmatrix},$$

and write the product determinant ΔD in the form

$$\Delta D \equiv \begin{vmatrix} 1 & u & u^2 & \ldots\ldots u^{\nu-1} & w \\ s_0 & s_1 & s_2 & \ldots\ldots s_{\nu-1} & \sigma_0 \\ s_1 & s_2 & s_3 & \ldots\ldots s_\nu & \sigma_1 \\ \ldots & \ldots & \ldots & \ldots\ldots & \ldots \\ s_{\nu-1} & s_\nu & s_{\nu+1} & \ldots\ldots s_{2\nu-2} & \sigma_{\nu-1} \end{vmatrix},$$

where $s_m = u_1^m + u_2^m + \ldots + u_\nu^m$, for $m = 0, 1, 2, \ldots\ldots, 2\nu - 2$, and

$\sigma_m = u_1^m\xi + u_2^m\epsilon\xi + \ldots + u_\nu^m\epsilon^{\nu-1}\xi$, for $m = 0, 1, 2, \ldots\ldots, \nu - 1$.

It appears that each s_m is expressible rationally in terms of the coefficients $g_1, g_2, \ldots\ldots, g_\nu$ appearing in the expression for G and, hence, is a rational function of z. In particular, $s_0 = \nu$. The functions σ_m are power series in ξ. Indeed, they are power series in z but need not be rational in z. In particular, $\sigma_0 = 0$. Since Δ is different from zero, the equation $\Delta D = 0$ implies the same function w as the earlier equation $D = 0$. Indeed, this fact can be seen from the form of the determinant ΔD itself. For, replacing u by u_1 and w by ξ, it appears that ΔD takes the value zero, seeing that it is then the sum of $\underline{|\nu}$ products of the form

$$\pm u_p^0 u_q^1 \ldots u_t^{\nu-1} \begin{vmatrix} 1 & u_1 & u_1^2 & \ldots\ldots u_1^{\nu-1} & \xi \\ 1 & u_p & u_p^2 & \ldots\ldots u_p^{\nu-1} & \epsilon^{p-1}\xi \\ 1 & u_q & u_q^2 & \ldots\ldots u_q^{\nu-1} & \epsilon^{q-1}\xi \\ \ldots & \ldots & \ldots & \ldots\ldots & \ldots \\ 1 & u_t & u_t^2 & \ldots\ldots u_t^{\nu-1} & \epsilon^{t-1}\xi \end{vmatrix},$$

in which $p, q, \ldots\ldots, t$ represents a permutation of $1, 2, \ldots\ldots, \nu$. Since the second factor of each of these products is zero, it follows that ΔD is also zero. The same conclusion follows if u is replaced by u_i and w by $\epsilon^{i-1}\xi$, for the remaining values of i. In other words, we see that w takes the values $\xi, \epsilon\xi, \ldots\ldots, \epsilon^{\nu-1}\xi$ as u takes the values $u_1, u_2, \ldots, u_\nu$. The difficulty arising from the fact that w need not involve z rationally can be dealt with by replacing $u_1, u_2, \ldots\ldots, u_\nu$, wherever they appear in D or Δ, by $\bar{u}_1, \bar{u}_2, \ldots\ldots, \bar{u}_\nu$, obtained from them by deleting all powers of ξ after a sufficiently high power. This has the effect of replacing the rational function s_m by a rational function $\bar{s}_m$ congruent to it up to a sufficiently high power of ξ, and likewise replacing the function σ_m by a rational function $\bar{\sigma}_m$ congruent to it up to a sufficiently high power of ξ. The function w resulting from this deletion has the properties required. That is, it is a rational function of (z, u) and has order 1 in z for the cycle $G = 0$.

Having thus seen that there exists a function $w = \rho(z, u)$, rational in (z, u), with orders $\frac{1}{\nu}$ in z for the branches of $G = 0$, or total order 1 in z for the cycle $G = 0$, we can employ it to construct a rational function of (z, u) with orders $\tau = \frac{\kappa}{\nu}$ in z for the branches of $G = 0$, or total order $\tau\nu = \kappa$ in z for the cycle $G = 0$, it being understood that κ is an integer. A satisfactory rational function of (z, u) is supplied by w^κ, understood to be written in reduced form relative to $G = 0$.

We say that a function-basis

$$(W) \equiv (W^{(1)}, W^{(2)}, \ldots\ldots, W^{(\nu)})$$

and an order-basis

$$(\tau) \equiv \left(\frac{\kappa}{\nu}\right)$$

in z for $G = 0$ are equivalent, provided every rational function of (z, u) built on either basis is built on the other as well. The rational functions of (z, u) built on (W) are all of the form

$$P^{(1)} W^{(1)} + P^{(2)} W^{(2)} + \ldots\ldots + P^{(\nu)} W^{(\nu)},$$

where $P^{(1)}, P^{(2)}, \ldots\ldots, P^{(\nu)}$ are rational in z and have orders zero or greater at the point $z = 0$. The rational functions of (z, u) built on (τ) all have total orders equal to or greater than κ in z for $G = 0$. It is plain that a function-basis (W) equivalent to the order-basis $(\tau) \equiv \left(\frac{\kappa}{\nu}\right)$ in z for $G = 0$ is supplied by

$$(W) \equiv (w^\kappa, w^{\kappa+1}, \ldots\ldots, w^{\kappa+\nu-1}).$$

The matrix M formed from the coefficients of $u^{\nu-1}, u^{\nu-2}, \ldots, 1$ appearing in the functions $W^{(1)}, W^{(2)}, \ldots\ldots, W^{(\nu)}$ of a function-basis

$$(W) \equiv (W^{(1)}, W^{(2)}, \ldots\ldots, W^{(\nu)})$$

in z for $G = 0$, constitutes a convenient representation of the function-basis. The rows of M refer to the functions $W^{(1)}, W^{(2)}, \ldots\ldots, W^{(\nu)}$ in order and the columns to $u^{\nu-1}, u^{\nu-2}, \ldots\ldots, 1$ in order.

5. *Function-basis for a combination of cycles equivalent to an assigned order-basis*

Let the fundamental equation $F = 0$ break up into r cycles at $z = 0$

$$G_1 = 0, G_2 = 0, \ldots\ldots, G_r = 0,$$

of degrees $\nu_1, \nu_2, \ldots\ldots, \nu_r$ in u, where, of course, $\nu_1 + \nu_2 + \ldots + \nu_r = N$. Let a total order-basis relative to $z = 0$ for the fundamental equation $F = 0$ be denoted by

$$(\tau) \equiv (\tau_1, \tau_2, \ldots\ldots, \tau_r),$$

where

$$(\tau_1) \equiv \left(\frac{\kappa_1}{\nu_1}\right), \quad (\tau_2) \equiv \left(\frac{\kappa_2}{\nu_2}\right), \ldots\ldots, \quad (\tau_r) \equiv \left(\frac{\kappa_r}{\nu_r}\right)$$

are partial order-bases relative to $z = 0$ for the cycles $G_1 = 0$, $G_2 = 0$, $\ldots\ldots$, $G_r = 0$ respectively. It will be found that the construction of a total function-basis

$$(W) \equiv (W^{(1)}, W^{(2)}, \ldots\ldots, W^{(N)})$$

equivalent to the total order-basis

$$(\tau) \equiv (\tau_1, \tau_2, \ldots\ldots, \tau_r)$$

involves the uses of partial order-bases

$$(\tau_1 - \omega_1), (\tau_2 - \omega_2), \ldots\ldots, (\tau_r - \omega_r)$$

rather than the partial order-bases

$$(\tau_1), (\tau_2), \ldots\ldots, (\tau_r),$$

where $\omega_1, \omega_2, \ldots\ldots, \omega_r$ are the branch orders of $\dfrac{F}{G_1}, \dfrac{F}{G_2}, \ldots\ldots, \dfrac{F}{G_r}$ at $z = 0$ for the cycles $G_1 = 0$, $G_2 = 0, \ldots\ldots, G_r = 0$ respectively.

An important special total order-basis at $z = 0$ for $F = 0$ is supplied by the branch orders of $\dfrac{\partial F}{\partial u}$ for $F = 0$ and is denoted by

$$(\mu) \equiv (\mu_1, \mu_2, \ldots\ldots, \mu_r).$$

It is clear that

$$(\mu_1 - \omega_1), (\mu_2 - \omega_2), \ldots\ldots, (\mu_r - \omega_r)$$

are special partial order-bases at $z = 0$ for $G_1 = 0, G_2 = 0, \ldots\ldots, G_r = 0$ respectively, supplied by the branch orders at $z = 0$ of

$$\frac{\partial G_1}{\partial u} \text{ for } G_1 = 0, \quad \frac{\partial G_2}{\partial u} \text{ for } G_2 = 0, \ldots\ldots, \quad \frac{\partial G_r}{\partial u} \text{ for } G_r = 0$$

respectively.

For the cycle $G_i = 0$ at $z = 0$, let

$$(W_i) \equiv (W_i^{(1)}, W_i^{(2)}, \ldots\ldots, W_i^{(\nu_i)})$$

by a partial function-basis equivalent to the partial order-basis

$$(\tau_i - \omega_i) \equiv \left(\frac{\kappa_i}{\nu_i}\right),$$

all the functions being of reduced form relative to $G_i = 0$, the whole applying for the individual cases $i = 1, 2, \ldots, r$. We wish to show how to employ these r partial function-bases

$$(W_1), (W_2), \ldots\ldots, (W_r)$$

for $G_1 = 0, G_2 = 0, \ldots\ldots, G_r = 0$ equivalent to the r partial order-bases

$$(\tau_1 - \omega_1), (\tau_2 - \omega_2), \ldots\ldots, (\tau_r - \omega_r)$$

for $G_1 = 0$, $G_2 = 0, \ldots\ldots, G_r = 0$ to construct a total function-basis

$$(W) \equiv (W^{(1)}, W^{(2)}, \ldots\ldots, W^{(N)})$$

for $F = 0$ equivalent to the total order-basis

$$(\tau) \equiv (\tau_1, \tau_2, \ldots\ldots, \tau_r)$$

for $F = 0$.

Another statement of the problem can be given in terms of matrices. Being supplied with partial matrices

$$M_1, M_2, \ldots\ldots, M_r$$

associated with partial function-bases

$$(W_1), (W_2), \ldots\ldots, (W_r)$$

equivalent to the partial order-bases

$$(\tau_1 - \omega_1), (\tau_2 - \omega_2), \ldots\ldots, (\tau_r - \omega_r),$$

we wish to show how to employ them to construct a total matrix M associated with the total function-basis (W) equivalent to the total order-basis (τ).

The statement of the result given in terms of function-bases is

$$(W) \equiv \left(\frac{F}{G_1}(W_1), \frac{F}{G_2}(W_2), \ldots\ldots, \frac{F}{G_r}(W_r) \right),$$

which is an abbreviated notation for

$$(W) \equiv \left(\frac{F}{G_1} W_1^{(1)}, \ldots, \frac{F}{G_1} W_1^{(\nu_1)}, \frac{F}{G_2} W_2^{(1)}, \ldots, \frac{F}{G_2} W_2^{(\nu_2)}, \right.$$
$$\left. \ldots\ldots, \frac{F}{G_r} W_r^{(1)}, \ldots, \frac{F}{G_r} W_r^{(\nu_r)} \right).$$

In terms of matrices, this result is

$$M \equiv \left| \begin{matrix} M_1 & & & & & & \\ & M_2 & & & & & \\ & & \cdot & & & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & & & & \cdot & \\ & & & & & & M_r \end{matrix} \right| R(G_1, G_2, \ldots\ldots, G_r),$$

as follows at once from the form stated for (W), on effecting the required multiplication of matrices. It remains, therefore, to show that (W) as defined is equivalent to (τ) as given. In the first place, any rational function of (z, u) built on the total function-basis

$$(W) \equiv \left(\frac{F}{G_1}(W_1), \frac{F}{G_2}(W_2), \ldots\ldots, \frac{F}{G_r}(W_r)\right)$$

is built on the total order-basis

$$(\tau) \equiv (\tau_1, \tau_2, \ldots\ldots, \tau_r),$$

seeing that this is true of the N particular rational functions of (z, u)

$$\frac{F}{G_i} W_i^{(j)}, \ (i = 1, \ldots, r\ ;\ j = 1, \ldots, \nu_i),$$

constituting (W). Indeed, the particular function $\frac{F}{G_i} W_i^{(j)}$ has branch orders equal to or more than $\omega_i + \tau_i - \omega_i = \tau_i$ for the cycle $G_i = 0$ and infinite branch orders for the remaining cycles. In the second place, any rational function of (z, u) built on the total order-basis

$$(\tau) \equiv (\tau_1, \tau_2, \ldots\ldots, \tau_r)$$

is built on the total function-basis

$$(W) \equiv \left(\frac{F}{G_1}(W_1), \frac{F}{G_2}(W_2), \ldots\ldots, \frac{F}{G_r}(W_r)\right)$$

as well. For, let $\rho(z, u)$ denote any assigned rational function of (z, u) built on the total order-basis (τ). Writing $\rho(z, u)$ in the Lagrange form corresponding to the cycle factorization

$$F \equiv G_1 G_2 \ldots\ G_r,$$

we have

$$\rho \equiv \frac{F}{G_1}\left(\frac{\rho G_1}{F}\right)_1 + \frac{F}{G_2}\left(\frac{\rho G_2}{F}\right)_2 + \ldots + \frac{F}{G_r}\left(\frac{\rho G_r}{F}\right)_r,$$

where $\left(\frac{\rho G_1}{F}\right)_1$, $\left(\frac{\rho G_2}{F}\right)_2, \ldots\ldots,$ $\left(\frac{\rho G_r}{F}\right)_r$ denote the reduced forms of $\frac{\rho G_1}{F}, \frac{\rho G_2}{F}, \ldots\ldots, \frac{\rho G_r}{F}$ relative to $G_1 = 0$, $G_2 = 0, \ldots\ldots, G_r = 0$ respectively. The portion

$$\frac{F}{G_i}\left(\frac{\rho G_i}{F}\right)_i$$

of the Lagrange form of ρ has branch orders equal to or more than τ_i for the cycle $G_i = 0$, and since the first factor $\frac{F}{G_i}$ has branch orders ω_i for $G_i = 0$, it follows that the second factor $\left(\frac{\rho G_i}{F}\right)_i$ has branch orders equal to or more than $\tau_i - \omega_i$ for $G_i = 0$ and, hence, is built on the partial function-basis (W_i) for $G_i = 0$. Of course, the portion $\frac{F}{G_i}\left(\frac{\rho G_i}{F}\right)_i$ has infinite branch orders for all cycles of $F = 0$ other than the particular cycle $G_i = 0$. In other words, $\frac{F}{G_i}\left(\frac{\rho G_i}{F}\right)_i$ is built on the portion of the function-basis (W) implied by $\frac{F}{G_i}(W_i)$. Seeing that this argument holds for $i = 1, 2, \ldots, r$, we have shown, therefore, that any rational function $\rho(z, u)$ built on the total order-basis (τ) is built on the total function-basis (W) as well, which, taken along with the earlier result in the opposite sense, concludes the proof of the equivalence of the total function-basis

$$(W) \equiv \left(\frac{F}{G_1}(W_1), \frac{F}{G_2}(W_2), \ldots, \frac{F}{G_r}(W_r)\right)$$

and the total order-basis

$$(\tau) \equiv (\tau_1, \tau_2, \ldots, \tau_r).$$

Briefly then, this means that the problem of constructing a total function-basis (W) equivalent to a total order-basis (τ) is reduced to the r separate problems, of simpler type, of setting up r partial function-bases $(W_1), (W_2), \ldots, (W_r)$ equivalent to r assigned partial order-bases $(\tau_1 - \omega_1), (\tau_2 - \omega_2), \ldots, (\tau_r - \omega_r)$, seeing that the act of combining them presents little formal difficulty, whether it be carried out on the actual functions involved or symbolically by the use of matrices.

6. *Properties of fundamental exponents*

The results to be discussed in this section refer to all the cycles of $F = 0$, taken at once. It must be realized that the total matrix M of a total function-basis

$$(W) \equiv (W^{(1)}, W^{(2)}, \ldots, W^{(N)})$$

equivalent to an assigned order-basis

$$(\tau) \equiv (\tau_1, \tau_2, \ldots, \tau_r)$$

at $z = 0$ has been defined relative to the N independent functions $u^{N-1}, u^{N-2}, \ldots\ldots, 1$. For that reason, we shall denote the total matrix M by

$$M \equiv M(u^{N-1}, u^{N-2}, \ldots, 1).$$

Total function-bases

$$(W) \equiv (W^{(1)}, W^{(2)}, \ldots, W^{(N)}),\ (\overline{W}) \equiv (\overline{W}^{(1)}, \overline{W}^{(2)}, \ldots, \overline{W}^{(N)})$$

are equivalent, according as their total matrices

$$M \equiv M(u^{N-1}, u^{N-2}, \ldots, 1),\quad \overline{M} \equiv \overline{M}(u^{N-1}, u^{N-2}, \ldots, 1)$$

are equivalent, or, what is the same thing, according as

$$M(u^{N-1}, u^{N-2}, \ldots, 1) \equiv \overline{M}(u^{N-1}, u^{N-2}, \ldots, 1)\,(a_{pq}),$$

where the matrix (a_{pq}) has order zero at $z = 0$, with its individual terms rational functions of z, of order zero or more at $z = 0$.

The determinants of equivalent total matrices

$$M(u^{N-1}, u^{N-2}, \ldots, 1),\ \overline{M}(u^{N-1}, u^{N-2}, \ldots, 1)$$

are of the same orders

$$\theta_N(u^{N-1}, u^{N-2}, \ldots, 1),\ \bar{\theta}_N(u^{N-1}, u^{N-2}, \ldots, 1)$$

at $z = 0$. But, the corresponding orders

$$\theta_i(u^{N-1}, u^{N-2}, \ldots, 1),\ \bar{\theta}_i(u^{N-1}, u^{N-2}, \ldots, 1),\ (i = 1, \ldots, N-1),$$

are also equal, where these denote the least orders of any of the $\binom{N}{i}^2$ sub-determinants of i rows and columns found in the matrices

$$M(u^{N-1}, u^{N-2}, \ldots, 1),\ \overline{M}(u^{N-1}, u^{N-2}, \ldots, 1)$$

respectively. For, any such sub-determinant of M is the sum of products of such sub-determinants of $\overline{M}$ and rational functions of z of order zero or more at $z = 0$, which means, therefore, that θ_i is not less than $\bar{\theta}_i$. The inverse total matrix identity

$$\overline{M} \equiv M(a_{pq})^{-1}$$

may be employed in like manner to show that $\bar{\theta}_i$ is not less than θ_i. In other words, $\theta_i = \bar{\theta}_i$.

Among total matrices $\overline{M}(u^{N-1}, u^{N-2}, \ldots, 1)$ equivalent to $M(u^{N-1}, u^{N-2}, \ldots, 1)$ those that are referred to as normalized are particularly important. Their principal diagonal terms are

$$z^{-\alpha_{N-1}},\ z^{-\alpha_{N-2}}, \ldots\ldots, z^{-\alpha_0},$$

and the terms to the left of the principal diagonal are all zero. The numbers (α), given by

$$(\alpha) \equiv (\alpha_{N-1}, \alpha_{N-2}, \ldots, \alpha_0),$$

are called the fundamental exponents associated with $(u^{N-1}, u^{N-2}, \ldots, 1)$ and the total order-basis $(\tau) \equiv (\tau_1, \tau_2, \ldots, \tau_r)$ at $z = 0$. The fundamental exponents (α) can, however, be found without carrying out the normalization of $M(u^{N-1}, u^{N-2}, \ldots, 1)$. This we shall now indicate for the case that u has all its orders zero or more at $z = 0$. Each diagonal term in a normalised matrix equivalent to $M(u^{N-1}, u^{N-2}, \ldots, 1)$ has the least order in z of all terms in the same row. For this reason and since $\alpha_{N-1}, \alpha_{N-2}, \ldots, \alpha_0$ are in monotone decreasing order, we see that

$$(\alpha_{N-1} + \alpha_{N-2} + \ldots + \alpha_{N-i}) = -\theta_i,\ (i = N-1, N-2, \ldots, 1),$$

where it must be understood that the sets (α), (θ) are both associated with $(u^{N-1}, u^{N-2}, \ldots\ldots, 1)$. On the other hand, if u has some of its orders negative at $z = 0$, we shall employ an integer h so that $v = uz^h$ has all its orders zero or more at $z = 0$. Where the fundamental exponents associated with (τ) and $(u^{N-1}, u^{N-2}, \ldots\ldots, 1)$ are

$$(\alpha_{N-1}, \alpha_{N-2}, \ldots\ldots, \alpha_0),$$

those associated with (τ) and $(v^{N-1}, v^{N-2}, \ldots\ldots, 1)$ are

$$(\alpha_{N-1} + (N-1)h,\ \alpha_{N-2} + (N-2)h, \ldots\ldots, \alpha_0).$$

If the columns of $M(u^{N-1}, u^{N-2}, \ldots\ldots, 1)$ are divided in order by $z^{\alpha_{N-1}}, z^{\alpha_{N-2}}, \ldots\ldots, z^{\alpha_0}$, we obtain thereby $M(v^{N-1}, v^{N-2}, \ldots\ldots, 1)$, from which the numbers (θ) can be obtained. We have, therefore,

$$(\alpha_{N-1} + \alpha_{N-2} + \ldots + \alpha_{N-i}) + \left(N - \frac{i+1}{2}\right) ih = -\bar{\theta}_i,$$

$$(i = N-1, N-2, \ldots, 1),$$

where the sets (α), $(\bar{\theta})$ are associated with $(u^{N-1}, u^{N-2}, \ldots\ldots, 1)$, $(v^{N-1}, v^{N-2}, \ldots\ldots, 1)$ respectively. In either case, the numbers (α) can be found in terms of known quantities.

The fundamental exponents (α) associated with (τ) and $(u^{N-1}, u^{N-2}, \ldots\ldots, 1)$ satisfy the relation (2)

$$\Sigma\alpha + \Sigma\tau\nu = \tfrac{1}{2}\Sigma\left(\mu - 1 + \frac{1}{\nu}\right)\nu.$$

It should be realized that the relation continues to hold if stated in terms of $v = uz^h$ rather than u. This is due to the compensating facts that the set (α) converts into

$$(\alpha_{N-1} + (N-1)h,\ \alpha_{N-2} + (N-2)h, \ldots\ldots, \alpha_0)$$

and the order-basis (μ) into $(\mu + (N-1)h)$.

We propose to examine what happens to the fundamental exponents (α) when any particular one of the order numbers in (τ) is increased by the minimum amount, that is to say when a single τ_i is increased by the corresponding $\frac{1}{\nu_i}$. Already it is clear that none of the numbers in (α) increases. But, since their sum decreases by 1, it follows that only one of them actually decreases, and that indeed by 1. It is proposed to indicate which of them it is that suffers this decrease of 1. The discussion need be given only for the case that μ has none but zero or positive orders at $z = 0$, seeing that the remaining case can be reduced to it by making use of $v = uz^h$, for h sufficiently large.

Taking $(W) \equiv (W^{(1)}, W^{(2)}, \ldots\ldots, W^{(N)})$ to be a total function-basis equivalent to a total order-basis $(\tau) \equiv (\tau_1, \tau_2, \ldots\ldots, \tau_r)$ at $z = 0$, there is no loss of generality in supposing that only one of the N functions, say $W^{(j)}$, in the total function-basis (W), has the precise order τ_i for the branches of the cycle $G_i = 0$. We shall now prove the lemma that the function-basis (W') obtained from (W) by replacing $W^{(j)}$ by $z\,W^{(j)}$ is equivalent to the order-basis (τ') obtained from (τ) by replacing τ_i by $\tau_i + \dfrac{1}{\nu_i}$. For one thing, all rational functions $\rho'(z, u)$ built on (W') are certainly built on (τ') as well. We have still to see, however, that all rational functions $\rho'(z, u)$ built on (τ') are built on (W') also. The general rational function $\rho(z, u)$ built on (τ) is of the form

$$\rho \equiv P^{(1)}W^{(1)} + P^{(2)}W^{(2)} + \ldots . + P^{(N)}W^{(N)},$$

where $P^{(1)}, P^{(2)}, \ldots . , P^{(N)}$ are rational functions of z, with orders zero or more at $z = 0$. From this we obtain the general rational function $\rho'(z, u)$ built on (τ') by making it impossible for the exact order τ_i to be taken by it for branches of the cycle $G_i = 0$. This is done by requiring the constant coefficient in $P^{(j)}$ to be zero, which converts $P^{(j)}$ into a function with z as factor. In other words, ρ' is built on (W'). This completes the proof of the equivalence of (τ') and (W') at $z = 0$.

Beginning with the equivalent total order-basis (τ) and function-basis (W), with which are associated the sets (α) and (θ), and passing to equivalent total order-basis (τ') and function-basis (W'), with which are associated the sets (α') and (θ'), we are now in a position to outline a procedure for discovering which of the N differences implied by $(\alpha' - \alpha)$ is -1. On multiplying the elements in row j of the matrix $M(u^{N-1}, u^{N-2}, \ldots . , 1)$ associated with (τ) by z, we obtain a matrix $M'(u^{N-1}, u^{N-2}, \ldots . , 1)$ associated with (τ'). In particular, therefore, $\theta'_N = 1 + \theta_N$. With respect to the correspondence between θ'_{N-1} and θ_{N-1} there are two possibilities, namely

$$\text{(i)}\ \theta'_{N-1} = \theta_{N-1};\quad \text{(ii)}\ \theta'_{N-1} = 1 + \theta_{N-1}.$$

Case (i) arises when, if row j be deleted from M, there exists in the remainder a sub-determinant containing $N - 1$ rows and columns and having order θ_{N-1} in z. On the other hand, case (ii) arises when no such sub-determinant exists. These statements are evident, and they provide the means of distinguishing between cases (i) and (ii). In case (i), therefore, the information sought has been obtained, since

$$\alpha'_0 = -1 + \alpha_0,\ \alpha'_1 = \alpha_1,\ \alpha'_2 = \alpha_2, \ldots . , \alpha'_{N-1} = \alpha_{N-1}.$$

In case (ii), however, we have found out merely that $\alpha'_0 = \alpha_0$, and so the process has to be continued to the next step. Cases (i) and (ii) of this next step are obtained from the original forms already given, by replacing $N-1$ by $N-2$. In case (i) of this next step, complete information has been obtained, since, of course,

$$a_1' = -1 + a_1,\ a_2' = a_2, \ldots\ldots\ldots\ldots, a_{N-1}' = a_{N-1}$$

is the new information to be associated with $a_0' = a_0$ of the original case (ii). In case (ii) of this next step, the new information obtained is merely $a_1' = a_1$, to be associated with $a_0' = a_0$ of the original case (ii), and still another step must be taken. The process may be continued in this way until the desired information is obtained.

The cases that apply may be listed as in the schedule involving N rows and $N-1$ columns:

(i)
(ii) (i)
(ii) (ii) (i)

(ii) (ii) (ii) .... (i)
(ii) (ii) (ii) .... (ii)

Each row states whether (i) or (ii) happens, up to the necessary stage of application. These rows correspond to the results:

$$\begin{aligned} a_0' &= -1 + a_0; \\ a_1' &= -1 + a_1; \\ a_2' &= -1 + a_2; \\ & \\ a_{N-2}' &= -1 + a_{N-2}; \\ a_{N-1}' &= -1 + a_{N-1}; \end{aligned}$$

which register by rows the particular number in the set $(a' - a)$ which is actually -1, leaving the remaining numbers in the set zero.

If a total order-basis (τ) is replaced by any other total order-basis (τ'), a corresponding change of fundamental exponents from (a) to (a') results. What these changes are may be found by applying the preceding unit result either up or down and as often as necessary, observing always the condition for applying it that only one of the functions in the total function-basis takes for the particular cycle at which the order number is stepped up by the minimum amount the exact order appearing in the equivalent total order-basis.

Given a total order-basis $((\tau))$ made up of individual total order-bases for all finite points, it has been shown (3) how to combine the corresponding individual total function-bases to form a single total function-basis $((W))$, with matrix M, applicable to all finite points at once.

REFERENCES

(1) On the factorization of certain functions rational and integral in one of several variables. Proc. Lond. Math. Soc., **24**, ser. **2**: 339-46.

(2) The theory of algebraic functions based on a study of fundamental exponents. Scripta Mathematica, **5**: 187-94, 239-45. 1938.

(3) Integral bases for an algebraic function field. Jour. Lond. Math. Soc., **4**: 13-17.

TRANSACTIONS OF

THE ROYAL SOCIETY
OF CANADA

SECTION III
CHEMICAL, MATHEMATICAL, AND PHYSICAL SCIENCES

THIRD SERIES—VOLUME XXXIX—SECTION III
MAY, 1945

56668

Ottawa
The Royal Society of Canada
1945

TABLE OF CONTENTS

A Six-Body Problem. By DANIEL BUCHANAN, F.R.S.C. 1

An Elementary Proof of a Theorem in Geophysics. By T. H. MATTHEWS 21

The Absorption Spectrum of Liquid Oxygen. By H. D. SMITH and N. BARTON 25

Wave-Impedances and the Effective Cross-Sections of Antennas. By W. H. WATSON, F.R.S.C. 33

Transactions of the Royal Society of Canada
SECTION III
SERIES III MAY, 1945 VOL. XXXIX

A SIX-BODY PROBLEM

By DANIEL BUCHANAN, F.R.S.C.

1. INTRODUCTION

This paper deals with periodic and asymptotic orbits in a particular problem of six bodies. Five of the bodies are of equal finite masses, are relatively fixed at the vertices of a regular pentagon, and revolve in circles about their common centre of gravity with uniform angular velocity. The sixth body is infinitesimal and moves in the vicinity of the equilibrium points or points of libration of the finite masses. The law of attraction is that of the inverse square.

In a recent paper, Hinrichsen (1913) considered the number of libration points for n bodies in which $n-1$ of them are finite and are relatively fixed at the vertices of a regular $(n-1)$-sided polygon rotating about the centre of gravity with uniform angular velocity. He showed that there are $3n-2$ libration points, all lying on axes of symmetry, and obtained their approximate position for the cases when $n = 3, 4, 5$. Periodic and asymptotic orbits near the libration points for $n = 5$ were obtained by the author in a recent paper (Buchanan, 1944).

The problem discussed in the present paper is to locate the equilibrium points for $n = 6$ and to determine the periodic and asymptotic orbits in their vicinity.

Two- and three-dimensional periodic orbits and two-dimensional asymptotic orbits have been obtained. The periodic orbits are similar to the orbits of Class A and Class B of Moulton's Oscillating Satellites (1920). The asymptotic orbits are akin to those obtained by the author (Buchanan, 1919) for the Asymptotic Satellites.

2. THE DIFFERENTIAL EQUATIONS

Following Hinrichsen, we choose a system of rectangular axes, xy, rotating uniformly about the centre of gravity. The plane of rotation is taken as the xy-plane and the co-ordinates of the finite masses are chosen as in Fig. 1. The sum of the masses is taken as the unit of mass and the unit of distance is selected as the distance from any finite mass to the centre of gravity of the remaining four finite masses. Thus the distance from the origin to any vertex of the pentagon is 4/5. The unit of time is so taken that the angular velocity of rotation is unity.

With the units and axes selected as above, the co-ordinates of the finite masses are

$$(x_k, y_k) = (.8 \cos (2k - 1)36^\circ, .8 \sin (2k - 1)36^\circ). \\ (k = 1, \ldots, 5). \qquad (1)$$

Let the co-ordinates of the infinitesimal body be denoted by (x, y). Then the differential equations of its motion are

$$\left.\begin{aligned} &\frac{d^2x}{dt^2} - 2\frac{dy}{dt} = \Omega_x, \\ &\frac{d^2y}{dt^2} + 2\frac{dx}{dt} = \Omega_y, \\ &\Omega = \frac{1}{10}\sum_{i=1}^{5}\left(r_i^2 + \frac{2}{r_i}\right), \\ &r_i^2 = (x - x_i)^2 + (y - y_i)^2, \\ &i = 1, \ldots, 5. \end{aligned}\right\} \qquad (2)$$

3. THE LIBRATION POINTS

The libration points, as previously stated, are located on axes of symmetry. There are sixteen $(3n-2)$ libration points and five axes of symmetry. Hence, besides the origin, there are three libration points on each axis of symmetry. The axes of symmetry are the lines through the vertices of the pentagon and perpendicular to the opposite sides. From the way in which the axes of reference were chosen, it is obvious that the x-axis is an axis of symmetry. The abscissas of the libration points on the x-axis are found by solving $\Omega_x = \Omega_y = y = 0$. The approximate co-ordinates are listed in Table I.

TABLE I

Libration Point	*Co-ordinates*
(*a*)	(−1.258,0)
(*o*)	(0,0)
(*b*)	(.822,0)
(*c*)	(.916,0)

The co-ordinates of the other points can be obtained by rotating the xy-axes through multiples of 72°. The libration points are marked in Fig. 1.

We shall consider in detail the points on the x-axis only.

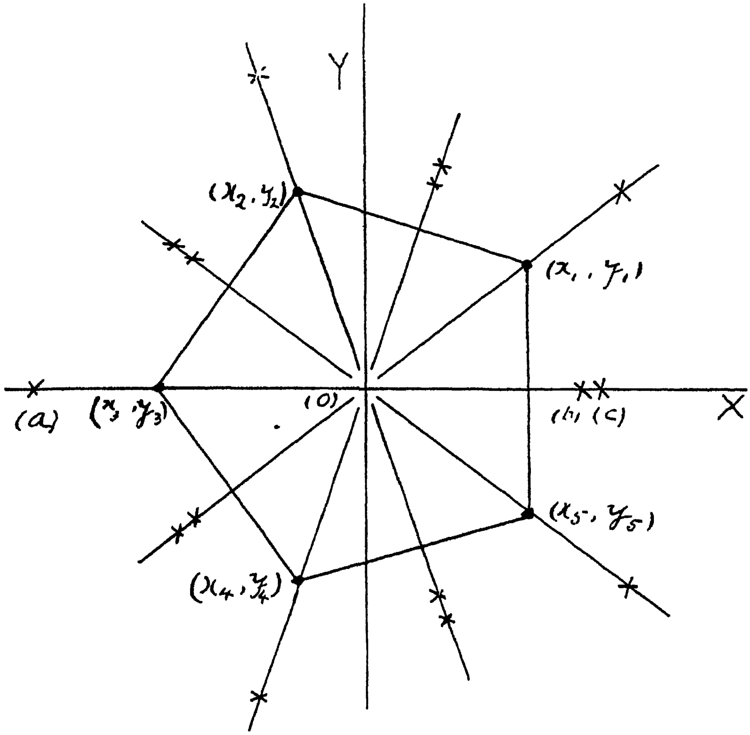


FIGURE 1

4. THE EQUATIONS OF DISPLACEMENT

Let us make in (2) the substitutions

$$x = \lambda + \epsilon u,\ y = \epsilon v,\ t - t_0 = (1 + \delta)\tau, \tag{3}$$

where λ denotes the abscissas of the libration points in Table I; u and v are new dependent variables; τ is a new independent variable; ϵ is a parameter and δ is a power series in ϵ vanishing with ϵ. Derivation with respect to τ will be denoted by the operator D. With the above substitutions, equations (2) become

$$\left.\begin{aligned}
&\epsilon \mathrm{D}^2 u - 2\epsilon (1 + \delta)\, \mathrm{D} v = \tfrac{1}{5} (1 + \delta)^2 \sum_{i=1}^{5}{}^{*} \left(1 - \frac{1}{\rho_i^3}\right)(\lambda - x_i + \epsilon u),\\
&\epsilon \mathrm{D}^2 v + 2\epsilon (1 + \delta)\, \mathrm{D} u = \tfrac{1}{5} (1 + \delta)^2 \sum_{i=1}^{5}{}^{*} \left(1 - \frac{1}{\rho_i^3}\right)(- y_i + \epsilon v),\\
&\rho_i^2 = (\lambda - x_i + \epsilon u)^2 + (- y_i + \epsilon v)^2.
\end{aligned}\right\} \tag{4}$$

The asterisk on the summation sign designates the omission of terms independent of ϵ. For sufficiently small numerical values of ϵ the terms in $1/\rho_i^3$ can be expanded into converging power series in ϵ. We thus obtain, after ϵ has been cancelled off,

$$\left.\begin{aligned}
[D^2 - d_1(1+\delta)^2]\,u - 2D\,(1+\delta)\,v &= (1+\delta)^2 \sum_{i=1}^{\infty} \sum_{j,k=0}^{i} U_{jk}^{(i)} u^j v^k \epsilon^i, \\
2D\,(1+\delta)\,u + [D^2 - d_2(1+\delta)^2]\,v &= (1+\delta)^2 \sum_{i=1}^{\infty} \sum_{j,k=0}^{i} V_{jk}^{(i)} u^j v^k \epsilon^i, \\
j + k &= i + 1,
\end{aligned}\right\} \quad (5)$$

where

$$d_1 = 1 - \tfrac{1}{5} \sum_{i=1}^{5} \left\{ \frac{1}{A_i^3} - \frac{3\,(\lambda - x_i)^2}{A_i^5} \right\},$$

$$d_2 = 1 - \tfrac{1}{5} \sum_{i=1}^{5} \left\{ \frac{1}{A_i^3} - \frac{3y_i^3}{A_i^5} \right\},$$

$$A_i^2 = (\lambda - x_i)^2 + y_i^2,$$

$$U_{20}^{(1)} = \tfrac{3}{10} \sum_{i=1}^{5} \left\{ \frac{3(\lambda - x_i)}{A_i^5} - \frac{5(\lambda - x_i)^3}{A_i^7} \right\},$$

$$U_{02}^{(1)} = \tfrac{3}{10} \sum_{i=1}^{5} \left\{ \frac{3(\lambda - x_i)}{A_i^5} - \frac{5(\lambda - x_i)y_i^2}{A_i^7} \right\},$$

$$V_{11}^{(1)} = 2U_{02}^{(1)}.$$

The remaining $U_{jk}^{(i)}$, $V_{jk}^{(i)}$ are functions of λ, x_i, y_i. In particular

$$U_{jj}^{(i)} = V_{j0}^{(i)} = V_{0j}^{(i)} = 0.$$

If we consider only the terms in (5) that are independent of ϵ, we obtain the equations of variation, viz.,

$$\begin{aligned}
(D^2 - d_1)\,u - 2Dv &= 0, \\
2Du + (D^2 - d_2)\,v &= 0.
\end{aligned} \qquad (6)$$

The indicial equation for the characteristic exponents is

$$D^4 + (4 - d_1 - d_2)\,D^2 + d_1 d_2 = 0.$$

The values of d_1, d_2, $U_{20}^{(1)}$, $U_{02}^{(1)}$ and of the characteristic equations and exponents for the libration points (*a*), (*o*), (*b*), (*c*) are listed in Table II.

The character of the solutions of the differential equations of motion (2) depends upon the solutions of the equations of variation. These latter solutions are known as generating solutions. If the characteristic exponents are purely imaginary the generating solutions are periodic and lead to periodic solutions of the equations of motion. If the characteristic exponents are real or complex, asymptotic solutions of the equations of motion will result. From the form of the characteristic exponents listed in Table II it follows that asymptotic orbits exist in the vicinity of each libration point (*a*), (*o*), (*b*), (*c*) and periodic orbits near (*a*) and (*b*) only.

TABLE II

CONSTANTS	(*a*) (−1.258,0)	(*o*) (0,0)	(*b*) (.882,0)	(*c*) (.916,0)
d_1	5.403	1.977	−.740	.590
d_2	−1.123	1.977	6.108	4.128
$U_{20}^{(1)}$	13.707	0	2.818	5.996
$U_{02}^{(1)}$	− 6.639	0	−11.210	−9.653
Characteristic Equations	$D^4 - .280D^2$ $-6.078 = 0$	$D^4 + .046D^2$ $+3.909 = 0$	$D^4 - 1.368D^2$ $-4.520 = 0$	$D^4 - .718D^2$ $+2.436 = 0$
Characteristic Exponents	± 1.615, $\pm 1.526i$	$\pm(.989 \pm i)$	$\pm 1.245i$, ± 1.708	$\pm(.980 \pm .775i)$

5. CONSTRUCTION OF TWO-DIMENSIONAL PERIODIC ORBITS

We propose to determine periodic solutions of equations (5), as power series in ϵ. Accordingly we put

$$u = \sum_{j=0}^{\infty} u_j \epsilon^j, \qquad v = \sum_{j=0}^{\infty} v_j \epsilon^j, \qquad \delta = \sum_{j=1}^{\infty} \delta_j \epsilon^j, \tag{7}$$

and substitute in (5). The resulting equations will be cited as (5′). By equating the coefficients of the various powers of ϵ in (5′), we obtain sets of differential equations that determine the various u_j and v_j. In the case of periodic orbits we proceed to show that the δ_j and the various constants of integration can be so determined that the u_j and v_j will be periodic and will satisfy certain assigned initial conditions.

Periodic Orbits near (*a*)

We shall consider in detail the construction of the two-dimensional periodic orbits near the point (*a*) and then list the results for the corresponding orbits near (*b*).

For initial values of u and v it will be supposed that $u = 1$, $v = 0$ at the initial time. There is no loss of generality in putting $u = 1$ as u carries the arbitrary factor ϵ in (3), nor in putting $v = 0$ as the initial time may be chosen as the instant when the infinitesimal crosses the x-axis. With these initial conditions it follows that

$$\left.\begin{aligned} &u_0(0) = 1, \quad u_j(0) = 0, \quad j = 1, 2, \ldots, \\ &v_j(0) = 0, \quad j = 0, 1, 2 \ldots. \end{aligned}\right\} \tag{8}$$

Step 1. Coefficients of ϵ^0 *in* (5′)

The differential equations at the first step are the same as the equations of variation and hence the periodic solutions that satisfy (8) are

$$u_0 = \cos\beta\tau,\ v_0 = -2.533\sin\beta\tau,\ \beta = 1.526.$$

The period of these solutions is $2\pi/\beta$.

Step 2. Coefficients of ϵ *in* (5′)

At the second step the differential equations are

$$\left.\begin{aligned} &(D^2 - d_1)u_1 - 2Dv_1 = U^{(2)} = \delta_1(3.075\cos\beta\tau) - \\ &14.445 + 28.152\cos 2\beta\tau), \\ &2Du_1 + (D^2 - d_2)v_1 = V^{(2)} = \delta_1(8.740\sin\beta\tau) + \\ &16.817\sin 2\beta\tau). \end{aligned}\right\} \tag{9}$$

The particular integrals of (9), expressed in symbolic form, are

$$\left.\begin{aligned} u_1 &= \frac{1}{\Delta}[(D^2 - d_2)U^{(2)} + 2DV^{(2)}], \\ v_1 &= \frac{1}{\Delta}[-2DU^{(2)} + (D^2 - d_1)V^{(2)}], \\ \Delta &= D^4 + (4 - d_1 - d_2)D^2 + d_1d_2. \end{aligned}\right\} \tag{10}$$

Since $\Delta = 0$ for $D^2 = -\beta^2$, the terms in $\cos\beta\tau$ and $\sin\beta\tau$ in (10) must vanish if u_1 and v_1 are to be periodic. As these terms carry the factor δ_1 we must put $\delta_1 = 0$ to obtain periodic solutions. The complete solutions, therefore, that are periodic and satisfy the initial conditions (8) are

$$\left.\begin{aligned} u_1 &= -1.115\cos\beta\tau + 2.669 - 1.554\cos 2\beta\tau, \\ v_1 &= 2.824\sin\beta\tau - .812\sin 2\beta\tau. \end{aligned}\right\} \tag{11}$$

Step 3. Coefficients of ϵ^2 *in* (5′)

From the integration of the equations at the third step we get an idea of the procedure at the general step. The differential equations at this step are

$$\left.\begin{aligned} (D^2 - d_1)u_2 - 2Dv_2 &= U^{(3)}, \\ 2Du_2 + (D^2 - d_2)v_2 &= V^{(3)}, \end{aligned}\right\} \tag{12}$$

$$U^{(3)} = \delta_2(3.075\cos\beta\tau) + \sum_{j=0}^{3} C_j^{(3)}\cos j\beta\tau,$$

$$V^{(3)} = \delta_2(8.740\sin\beta\tau) + \sum_{j=1}^{3} S_j^{(3)}\sin j\beta\tau,$$

where $C_j^{(3)}$, $S_j^{(3)}$ are known constants. The constant δ_2 enters (12) only as expressed.

The particular integrals of (12), expressed in symbolic form, are as in (10)

$$\left.\begin{aligned} u_2 &= \frac{1}{\Delta}\left[(D^2-d_2)\,U^{(3)}+2DV^{(3)}\right], \\ v_2 &= \frac{1}{\Delta}\left[-2DU^{(3)}+(D^2-d_1)V^{(3)}\right]. \end{aligned}\right\} \tag{13}$$

If u_2 and v_2 are to be periodic, the coefficients of $\cos\beta\tau$ and $\sin\beta\tau$ in (13) must vanish. Thus

$$\begin{aligned} -(\beta^2+d_2)\left\{3.075\,\delta_2+C_1^{(3)}\right\}+2\beta\left\{8.740\,\delta_2+S_1^{(3)}\right\} &= 0, \\ 2\beta\left\{3.075\,\delta_2+C_1^{(3)}\right\}-(\beta^2+d_1)\left\{8.740\,\delta_2+S_1^{(3)}\right\} &= 0. \end{aligned} \tag{14}$$

The functional determinant of the expressions in the brackets $\{\ \ \}$ vanishes and hence one equation of (14) is redundant. On solving either equation we obtain

$$\delta_2 = \frac{1.206\,C_1^{(3)} - 3.052\,S_1^{(3)}}{22.966}. \tag{15}$$

With δ_2 thus determined the periodic solutions satisfying the initial conditions $u_2(0)=v_2(0)=0$ are

$$u_2 = \sum_{j=0}^{3} c_j^{(3)}\cos j\beta\tau,\quad v_2 = \sum_{j=1}^{3} s_j^{(3)}\sin j\beta\tau,\quad \sum_{j=0}^{3} c_j^{(3)} = 0,$$

where the $c_j^{(3)}$ and $s_j^{(3)}$ are known constants.

Step $n+1$. Coefficients of ϵ^n in (5′)

Let us suppose that the solutions at all the preceding steps have been obtained and that for $k=1,\ldots,n-1$

$$\left.\begin{aligned} u_k &= \sum_{j=0}^{k+1} c_j^{(k+1)}\cos j\beta\tau, \\ v_k &= \sum_{j=1}^{k+1} s_j^{(k+1)}\sin j\beta\tau, \\ \delta_k &= \text{constant}. \end{aligned}\right\} \tag{16}$$

Then the differential equations at the step $n+1$ are

$$\begin{aligned} (D^2-d_1)\,u_n-2Dv_n &= U^{(n+1)}, \\ -2D\,u_n+(D^2-d_2)\,v_n &= V^{(n+1)}, \\ U^{(n+1)} &= \delta_n\,(3.075\cos\beta\tau)+\sum_{j=0}^{n+1} C_j^{(n+1)}\cos j\beta\tau, \\ V^{(n+1)} &= \delta_n\,(8.693\sin\beta\tau)+\sum_{j=1}^{n+1} S_j^{(n+1)}\sin j\beta\tau, \end{aligned}$$

where $C_j^{(n+1)}$, $S_j^{(n+1)}$ are constants and δ_n enters only as expressed.

On forming the symbolic expressions for the particular integrals we find that two equations in δ_n must be satisfied if the particular integrals are to be periodic. These equations are the same as (14) if δ_2, $C_1^{(3)}$, $S_1^{(3)}$ are replaced by δ_n, $C_1^{(n+1)}$, $S_1^{(n+1)}$ respectively. Hence both equations are satisfied by putting

$$\delta_n = \frac{1.206\, C_1^{(n+1)} - 3.052\, S_1^{(n+1)}}{22.966}.$$

The desired periodic solutions that satisfy the initial conditions are therefore the same as (16) for $k = n$.

This completes the integration for the periodic orbits near (*a*). So far as the computation has been carried out we have

$$x + 1.258 = \epsilon \cos \beta\tau + \epsilon^2(-1.115 \cos \beta\tau + 2.669 - 1.554 \cos 2\beta\tau) + \ldots,$$

$$y = -2.533\, \epsilon \sin \beta\tau + \epsilon^2(2.824 \sin \beta\tau - .812 \sin 2\beta\tau) + \ldots.$$

If we neglect the terms in ϵ^2 and higher powers we obtain the approximate orbit

$$\frac{(x + 1.258)^2}{\epsilon^2} + \frac{y^2}{(2.533\epsilon)^2} = 1.$$

This represents an ellipse with the centre at (*a*) and with the major axis parallel to the y-axis. The parameter ϵ denotes the initial displacement from (*a*) along the x-axis.

Periodic Orbits near (*b*)

The corresponding periodic orbits near the libration point (*b*) are

$$x - .822 = \epsilon \cos \beta\tau + \epsilon^2(.425 \cos \beta\tau + 1.103 - 1.528 \cos 2\beta\tau) + \ldots,$$

$$y = -.325\epsilon \sin \beta\tau - \epsilon^2 (.138 \sin \beta\tau + .235 \sin 2\beta\tau) + \ldots,$$

$$\beta = 1.245.$$

The approximate equation is the ellipse

$$\frac{(x - .822)^2}{\epsilon^2} + \frac{y^2}{(.325\epsilon)^2} = 1$$

having its centre at (*b*) and its major axis along the x-axis. The parameter ϵ denotes, as in the orbits near (*a*), the initial displacement from (*b*) along the x-axis.

6. ASYMPTOTIC ORBITS

According to the definition given by Poincaré (1892) a solution is asymptotic if it has the form $e^{at}\, P(t)$ where the real part of a is negative and P(t) is periodic. As the characteristic exponents at all the libration points have values such that their real parts are negative, asymptotic solutions exist for all the points.

Point (a)

To obtain the asymptotic orbits near point (a) we consider equations (5′) with $\delta = 0$ and $D = d/dt$. These equations will be cited as (5″).

Step 1. Coefficients of ϵ^0 *in* (5″)

The differential equations at the first step are the same as the equations of variation (6) and the desired solutions are

$$u_0 = Be^{-1.615t}, \quad v_0 = .865\,Be^{-1.615t},$$

where B is arbitrary. If we select the initial condition $u(0) = 1$, which may be done without loss of generality as u carries the factor ϵ in (3), we have

$$u_0(0) = 1,\ u_j(0) = 0,\ j = 1, 2, \ldots . \tag{17}$$

Then the asymptotic solutions at the first step are

$$u_0 = e^{-1.615t}, v_0 = .865e^{-1.615t}.$$

Step 2. Coefficients of ϵ *in* (5″)

On equating the coefficients of ϵ in (5″) we obtain

$$\left.\begin{aligned}(D^2 - d_1)\,u_1 - 2Dv_1 &= U^{(2)} = 8.736e^{-3.230t},\\ 2Du_1 + (D^2 - d_2)\,v_1 &= V^{(2)} = -\,11.490e^{-3.230t}.\end{aligned}\right\} \tag{18}$$

As $U^{(2)}$ and $V^{(2)}$ contain no terms in $e^{-1.615t}$ the particular integrals will contain no terms in t explicitly. The desired solutions satisfying the initial condition (17) are

$$\begin{aligned}u_1 &= -\,1.755(e^{-1.615t} - e^{-3.230t}),\\ v_1 &= -\,1.519e^{-1.615t} - .001e^{-3.230t}.\end{aligned}$$

Step $n + 1$. *Coefficients of* ϵ^n *in* (5″)

Let us suppose that $u_0, \ldots, u_{n-1}, v_0, \ldots, v_{n-1}$ have all been determined and that

$$\left.\begin{aligned}&u_j = \sum_{k=1}^{j+1} b_k^{(j+1)}\,e^{-k\alpha t},\ v_j = \sum_{k=1}^{j+1} c_k^{(j+1)}\,e^{-k\alpha t},\\ &\alpha = -1.615, \quad j = 0, \ldots, n-1,\\ &b_k^{(j+1)},\ c_k^{(j+1)} = \text{constants},\\ &\sum_{k=1}^{j+1} b_k^{(j+1)} = 0, \quad j > 0.\end{aligned}\right\} \tag{19}$$

Then the differential equations at the step $n+1$ are

$$\left.\begin{aligned} (D^2-d_1)\,u_n-2Dv_n &= U^{(n+1)},\\ 2Du_n+(D^2-d_2)\,v_n &= V^{(n+1)},\\ U^{(n+1)} &= \sum^{n+1} B_k^{(n+1)}\,e^{-kat},\\ V^{(n+1)} &= \sum_{k=2}^{n+1} C_k^{(n+1)}\,e^{-kat}, \end{aligned}\right\} \qquad (20)$$

where $B_k^{(n+1)}$, $C_k^{(n+1)}$ are known constants. As no terms in e^{-at} occur in the right members of (20) there will be no terms in t explicitly in the particular integrals.

On obtaining the complete solutions and selecting the constants of integration so as to satisfy the initial and asymptotic conditions it will be found that the desired solutions for u_n and v_n are the same as (19) for $j=n$.

Approximate equations for the asymptotic orbit are

$$\begin{aligned} x+1.258 &= \epsilon e^{-1.615t},\\ y &= .865\epsilon\, e^{-1.615t}. \end{aligned}$$

They represent a straight line through (a) having the slope .865. With respect to fixed axes through the origin the asymptotic orbit for a given value of ϵ is the dotted line approaching the path of (a) in Fig. 2.

Point (b)

The asymptotic orbits near point (b) are similar to those near point (a). So far as the computation was carried out the equations are

$$\begin{aligned} x-.822 &= \epsilon\, e^{-1.708t}+1.900\epsilon^2(e^{-1.708t}-e^{-3.416t})\ldots,\\ y &= -1.071\epsilon\, e^{-1.708t}-\epsilon^2(2.035\,e^{-1.708t}-1.982\,e^{-3.416t})\ldots. \end{aligned}$$

An approximate orbit is

$$x-.822=\epsilon\, e^{-1.708t},\quad y=-1.071\epsilon\, e^{-1.708t}.$$

It is a straight line through (b) having the slope -1.071. Its orbit with respect to fixed axes through the origin is the dotted line approaching the path of (b) in Fig. 2.

Point (c)

As the solutions for the asymptotic orbits near (c) and (o) are different from those near (a) and (b), we shall consider in detail the orbits near (c) and list the results for (o).

Step 1. Coefficients of ϵ^0 *in* (5″)

The differential equations at the first step are the same as the equations of variation. The generating solutions are therefore

$$\left.\begin{aligned} u_0 &= e^{-\alpha t}(B_1 \cos \beta t + B_2 \sin \beta t), \\ v_0 &= e^{-\alpha t}(-.305\,B_1 + .534\,B_2) \cos \beta t, \\ \alpha &= -.980, \ \beta = .775, \end{aligned}\right\} \qquad (21)$$

where B_1 and B_2 are arbitrary constants. As there are two arbitrary constants at each step of the integration we may impose two initial conditions.

Let it be supposed that $u(0) = 1$, $v(0) = 0$. Then

$$\begin{aligned} &u_0(0) = 1, \ u_j(0) = 0, j = 1, 2, \ldots, \\ &v_j(0) = 0, \ j = 0, 1, 2, \ldots. \end{aligned}$$

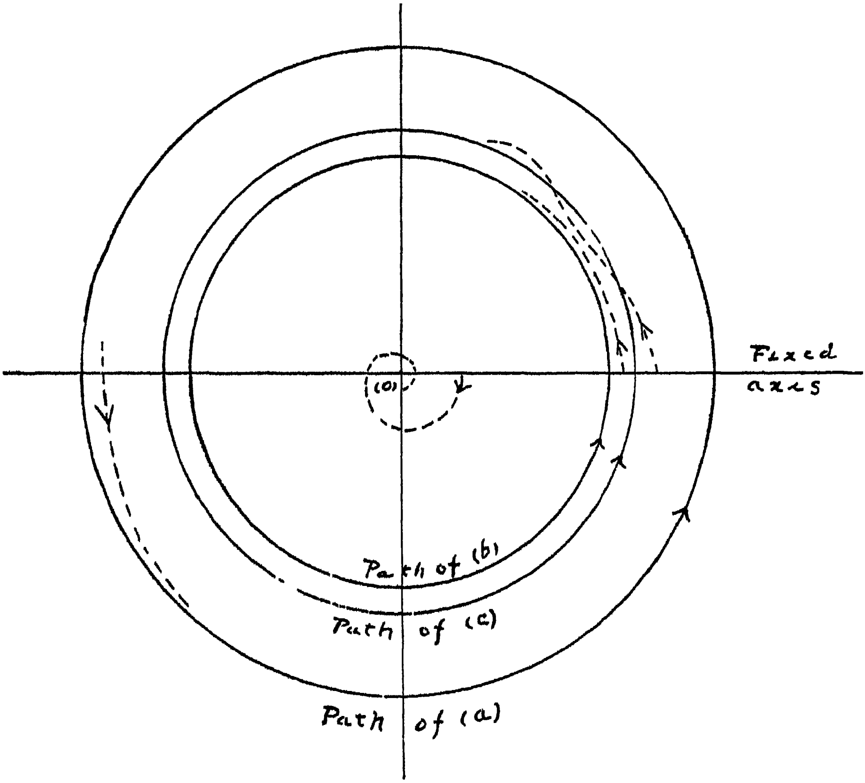


FIGURE 2

With these initial values the solutions at the first step are

$$\left.\begin{aligned} u_0 &= e^{-at}\,(\cos\beta t + .572\sin\beta t), \\ v_0 &= -e^{-at}\,(.708\sin\beta t). \end{aligned}\right\} \qquad (22)$$

Step 2. Coefficients of ϵ^2 *in* (5″)

The differential equations at the second step are

$$\left.\begin{aligned} &(\mathrm{D}^2 - d_1)u_1 - 2\,\mathrm{D}v_1 = \mathrm{U}^{(2)}, \\ &2\mathrm{D}\,u_1 + (\mathrm{D}^2 - d_2)\,v_1 = \mathrm{V}^{(2)}, \\ &\mathrm{U}^{(2)} = e^{-2at}\,(1.562 + 4.434\cos 2\beta t + 3.430\sin 2\beta t), \\ &\mathrm{V}^{(2)} = e^{-2at}\,(3.909 - 3.909\cos 2\beta t + 6.834\sin 2\beta t). \end{aligned}\right\} \qquad (23)$$

The particular integrals in symbolic form are the same as (10). On making use of the property

$$f(\mathrm{D})\,e^{at}\,\mathrm{U} = e^{at} f(\mathrm{D} + a)\,\mathrm{U},$$

$$\frac{1}{f(\mathrm{D})}\,e^{at}\,\mathrm{U} = e^{at}\,\frac{1}{f(\mathrm{D} + a)}\,u,$$

we note that the particular integrals will contain no terms in t explicitly. The complete asymptotic solutions at this step that satisfy the initial conditions are

$$\left.\begin{aligned} u_1 &= e^{-.980t}\,(1.161\cos\beta t - 3.805\sin\beta t) \\ &\quad - e^{-1.960t}\,(1.102 + .059\cos 2\beta t + .105\sin 2\beta t), \\ v_1 &= e^{-.980t}\,(-2.386\cos\beta t + .541\sin\beta t) \\ &\quad + e^{-1.960t}\,(1.305 + 1.081\cos 2\beta t + .148\sin 2\beta t). \end{aligned}\right\} \qquad (24)$$

Step n + 1. Coefficients of ϵ^n *in* (5″)

The integrations at this step are entirely similar to those at the second step. The desired solutions are

$$\begin{aligned} u_n &= \sum_{j=1}^{n+1} e^{-jat} \sum_{k=0}^{j} (\mathrm{B}_k^{(j)}\cos k\beta t + \mathrm{C}_k^{(j)}\sin k\beta t), \\ v_n &= \sum_{j=1}^{n+1} e^{-jat} \sum_{k=0}^{j} (\mathrm{D}_k^{(j)}\cos k\beta t + \mathrm{E}_k^{(j)}\sin k\beta t), \end{aligned} \qquad (25)$$

where $\mathrm{B}_k^{(j)}, \ldots, \mathrm{E}_k^{(j)}$ are constants. Further

$$\mathrm{B}_0^{(1)} = \mathrm{D}_0^{(1)} = 0, \quad \sum_{j=1}^{n+1}\sum_{k=0}^{j} \mathrm{B}_k^{(j)} = \sum_{j=1}^{n+1}\sum_{k=0}^{j} \mathrm{D}_k^{(j)} = 0.$$

An approximate orbit is obtained from the equations

$$\begin{aligned} x - .922 &= e^{-.980t}(\cos \beta t + .572 \sin \beta t)\,\epsilon, \\ y &= -\,e^{-.980t}(.708 \sin \beta t)\,\epsilon, \\ \beta &= .775. \end{aligned}$$

Table III lists values of $x - .922$ and y for certain values of k, where $t = k\pi/\beta$. Each value for the co-ordinates is to be multiplied by ϵ.

TABLE III

$4k$	$t = k\pi/\beta$	$X = x - .922$	y
0		1	0
1		.411	−.185
2		.0784	−.097
3		−.0154	−.0255
4		−.0188	0
5		−.00775	−.00350
6		.00148	.00183
7		.000290	.000479
8		.000355	0

With respect to fixed axes the asymptotic orbit is the dotted line approaching the path of (*c*) in Fig. 2. It crosses the path of (*c*) periodically and with diminishing amplitudes from the circle as is the case in damped vibrations.

Point (*o*)

The approximate asymptotic orbit near the origin is

$$\begin{aligned} x &= e^{-.980t}(\cos t)\,\epsilon, \\ y &= -e^{-.980t}(\sin t)\,\epsilon. \end{aligned}$$

Since $U^{(2)}_{20} = U^{(2)}_{02} = 0$ at the origin, the terms for the second step of the integration are zero. The orbit is a spiral about the origin as in Fig. 2 but with respect to the rotating xy-axes.

For all four points (*a*), (*o*), (*b*), (*c*) the parameter ϵ is arbitrary. From the way in which the initial conditions were chosen, ϵ is the scale factor denoting the initial displacement along the x-axis from a libration point.

7. THREE-DIMENSIONAL PERIODIC ORBITS

To obtain three-dimensional periodic orbits we choose a set of axes x, y, z in which the xy-plane and units are taken as in the previous work.

Let the co-ordinates of the infinitesimal be denoted now by (x, y, z). Then the differential equations of motion are

$$\left.\begin{aligned} &\frac{d^2x}{dt^2} - 2\frac{dy}{dt} = \Omega_x, \\ &\frac{d^2y}{dt^2} + 2\frac{dx}{dt} = \Omega_y, \\ &\qquad\frac{d^2z}{dt^2} = \Omega_z, \\ &\Omega = \tfrac{1}{10}\sum_{i=1}^{5}\left(r_i^2 + \frac{2}{r_i}\right), \\ &r_i^2 = (x - x_i)^2 + (y - y_i)^2 + z^2. \end{aligned}\right\} \qquad (26)$$

Let

$$u = \lambda + \epsilon u,\ y = \epsilon v,\ z = \epsilon w,\ t - t_0 = (1 + \delta)\tau, \qquad (27)$$

as in (3). Then the differential equations corresponding to (5) are

$$\left.\begin{aligned} &[D^2 - d_1(1+\delta)^2]\,u - 2D\,(1+\delta)\,v = \text{terms as in (5)} \\ &\qquad\qquad + (1+\delta)^2 U^{(2)}_{002}\,w^2\epsilon \\ &\qquad\qquad + \text{higher degree terms in } w^2, \\ &2D\,(1+\delta)\,u + [D^2 - d_2(1+\delta)^2]\,v = \text{terms as in (5)} \\ &\qquad\qquad + \text{higher degree terms in } w^2 \text{ but carrying} \\ &\qquad\qquad \text{the factor } v, \\ &[D^2 + \gamma^2(1+\delta)^2]\,w = (1+\delta)^2 W^{(2)}_{101}\,uw \\ &+ (1+\delta)^2\epsilon^2\,(W^{(3)}_{201}\,u^2w + W^{(3)}_{021}\,v^2w + W^{(3)}_{003}\,w^3) \\ &+ \text{higher degree terms in } w, \end{aligned}\right\} \qquad (28)$$

where

$$\begin{aligned} U^{(2)}_{002} &= \tfrac{3}{10}\sum_{i=1}^{5}\frac{\lambda - x_i}{A_i^5}, \quad \gamma = \text{constant}, \\ W^{(2)}_{101} &= 2\,U^{(2)}_{002}, \\ W^{(3)}_{201} &= \tfrac{3}{2}\sum_{i=1}^{5}\left[\frac{1}{5A_i^5} - \frac{(\lambda - x_i)^2}{A_i^7}\right], \\ W^{(3)}_{021} &= \tfrac{3}{2}\sum_{i=1}^{5}\left[\frac{1}{5A_i^5} - \frac{y_i^2}{A_i^7}\right], \\ W^{(3)}_{003} &= \tfrac{3}{10}\sum_{i=1}^{5}\frac{1}{A_i^5}. \end{aligned}$$

Now let

$$u = \sum_{j=0}^{\infty} u_j\epsilon^j,\ v = \sum_{j=0}^{\infty} v_j\epsilon^j,\ w = \sum_{j=0}^{\infty} w_j\epsilon^j,\ \delta = \sum_{j=1}^{\infty} \delta_j\epsilon^j. \qquad (29)$$

When these substitutions are made in (28) we obtain equations, cited as (28′), which define the various u_j, v_j, w_j. If we take only the linear terms in (28) we obtain the equations of variation, viz.,

$$\left.\begin{aligned}(D^2 - d_1)\,u - 2Dv &= 0,\\ 2Du + (D^2 - d_2)\,v &= 0,\\ (D^2 + \gamma^2)\,w &= 0.\end{aligned}\right\} \qquad (30)$$

The first two equations are the same as (6). It has been shown that these equations have periodic solutions for the points (a) and (b) only. Their periods are $P_1 = 2\pi/\beta$, where $\beta = 1.526$ for (a) and $\beta = 1.708$ for (b). The solution of the third equation of (30) is

$$w = C_1 \cos \gamma\tau + C_2 \sin \gamma\tau,$$

and its period is $P_2 = 2\pi/\gamma$, where γ has the values listed in Table IV. Hence there are three types of periodic solutions of (30) according as we select the period P_1 or P_2 if β and γ are incommensurable or the period P_3 which is a multiple of P_1 and P_2 when β and γ are commensurable.

For the period P_1 the generating solution of the w-equation must be zero and this will lead to the two-dimensional periodic orbits already determined. The case when β and γ are commensurable has been found too complicated and is left unsolved. The three-dimensional periodic orbits have the period $P_2 = 2\pi/\gamma$, β and γ being incommensurable. Their generating solutions are

$$u = v = 0,\ w = C_1 \cos \gamma\tau + C_2 \sin \gamma\tau. \qquad (31)$$

As two arbitrary constants occur in (31) we may impose two initial conditions. Since the infinitesimal must cross the xy-plane if its motion in three-dimensions is to be periodic, we may select the initial time as the instant when the infinitesimal is crossing the xy-plane. Thus we may take $w(0) = 0$ without loss of generality. Again, since w carries the arbitrary factor ϵ in (28) we may put $Dw(0) = 1$, likewise without loss of generality. When these initial values are imposed upon w we obtain

$$\left.\begin{aligned}&w_j(0) = 0, \quad j = 0, 1, 2, \ldots,\\ &D(w)_0(0) = 1, \quad Dw_j(0) = 0, j > 0\end{aligned}\right\} \qquad (32)$$

We shall now integrate equation (28′) subject to the initial and periodicity conditions selected.

Step 1. Coefficients of ϵ^0 *in* (28′)

The differential equations are the same as the equations of variation and the desired solutions are

$$u_0 = v_0 = 0, \quad w_0 = \frac{1}{\gamma} \sin \gamma\tau.$$

Step 2. Coefficients of ϵ *in* (28′)

The differential equations at this step are

$$\left.\begin{aligned} (D^2 - d_1)\, u_1 - 2Dv_1 &= U^{(2)}_{002}\, w_0^2, \\ 2Du_1 + (D^2 - d_2)\, v_1 &= 0, \\ (D^2 + \gamma^2)\, w_1 &= -\, 2\delta_1 w_0. \end{aligned}\right\} \qquad (33)$$

Since β and γ are incommensurable only the particular integrals of the first two equations of (33) may be retained. In the last equation we must have $\delta_1 = 0$ if w is to be periodic. The desired solutions at this step are therefore

$$\begin{aligned} w_1 &= c_0^{(1)} + c_2^{(1)} \cos 2\gamma\tau, \\ v_1 &= \qquad\quad s_2^{(1)} \sin 2\gamma\tau, \\ w_1 &= 0, \end{aligned}$$

where $c_0^{(1)}$, $c_2^{(1)}$, $s_2^{(1)}$ are known constants.

Step 3. Coefficients of ϵ^2 *in* (28′)

The differential equations at the third step are

$$\begin{aligned} (D^2 - d_1)u_2 - 2Dv_2 &= 0, \\ 2Du_2 + (D^2 - d_2)\, v_2 &= 0, \\ (D^2 + \gamma^2)w_2 &= -\left(\frac{2\delta_2}{\gamma} + G_1^{(2)}\right) \sin \gamma\tau + G_3^{(2)} \sin 3\gamma\tau, \end{aligned}$$

where $G_1^{(2)}$, $G_3^{(2)}$ are known constants. The desired solutions are

$$\begin{aligned} u_2 &= v_2 = 0, \\ w_2 &= g_1^{(2)} \sin \gamma\tau + g_3^{(2)} \sin 3\gamma\tau, \\ \delta_2 &= -\tfrac{1}{2}\gamma G_1^{(2)}. \end{aligned}$$

The integration of the higher terms proceeds as at the previous steps. It is found that the u's and v's with even subscripts and the w's and δ's with odd subscripts are all zero. The non-zero terms have the general form

$$\begin{aligned} u_{2k-1} &= \sum_{j=0}^{k} c_{2j}^{(2k-1)} \cos 2j\gamma\tau, \\ v_{2k-1} &= \sum_{j=1}^{k} s_{2j}^{(2k-1)} \sin 2j\gamma\tau, \\ w_{2k} &= \sum_{j=0}^{k} S_{2j+1}^{(2k+1)} \sin (2j+1)\gamma\tau, \end{aligned}$$

where $c_{2j}^{(2k-1)}$, $s_{2j}^{(2k-1)}$, are known constants.

The numerical values of various terms in the first three steps of the integration for the solutions near the four libration points are listed in Table IV.

TABLE IV

	(a)	(o)	(b)	(c)
$U_{002}^{(3)}$	-7.017	0	3.502	3.642
$W_{003}^{(3)}$	15.065	4.578	19.003	13.017
γ	1.146	$.976$	1.545	1.308
u_0	0	0	0	0
v_0	0	0	0	0
w_0	$1/\gamma \sin \gamma\tau$	$1/\gamma \sin \gamma\tau$	$1/\gamma \sin \gamma\tau$	$1/\gamma \sin \gamma\tau$
δ_1	0	0	0	0
u_1	$.494-.304 \cos 2\gamma\tau$	0	$.922+.115 \cos 2\gamma\tau$	$-1.805+.216 \cos 2\gamma\tau$
v_1	$.337 \sin 2\gamma\tau$	0	$-.046 \sin 2\gamma\tau$	$-.103 \sin 2\gamma\tau$
w_1	0	0	0	0
u_2	0	0	0	0
v_2	0	0	0	0
δ_2	$-.180$	1.890	2.622	-2.401
w_2	$-.244 \sin^3\gamma\tau$	$-.512 \sin^3\gamma\tau$	$-.216 \sin^3\gamma\tau$	$-.252 \sin^3\gamma\tau$

The approximate periodic orbits near each point are obtained by taking only the terms in ϵ and ϵ^2 in the solutions.

Point (a)

The equations of the approximate orbit near (a) are

$$\begin{aligned} x + 1.258 &= \epsilon^2 (.494 - .304 \cos 2\gamma\tau), \\ y &= \epsilon^2 (.337 \sin 2\gamma\tau), \\ z &= \epsilon/\gamma \sin \gamma\tau. \end{aligned}$$

The equations of the projections on the co-ordinate planes and the corresponding curves are as follows:

$$xy\text{-}plane \quad \frac{(x + 1.258 - .494\epsilon^2)^2}{(.304\epsilon^2)^2} + \frac{y^2}{(.337\epsilon^2)^2} = 1.$$

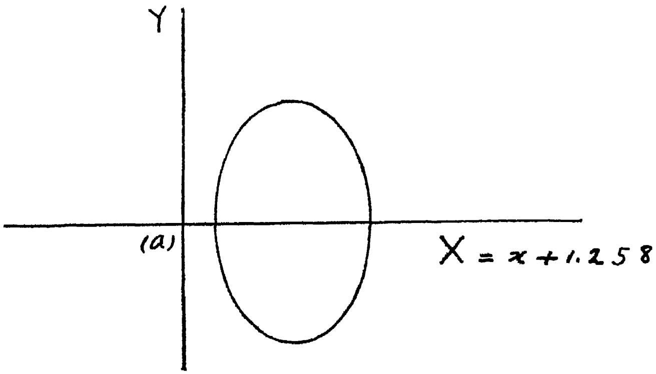


FIGURE 3

xz-plane

$$x + 1.258 - .190\epsilon^2 = .800z^2.$$

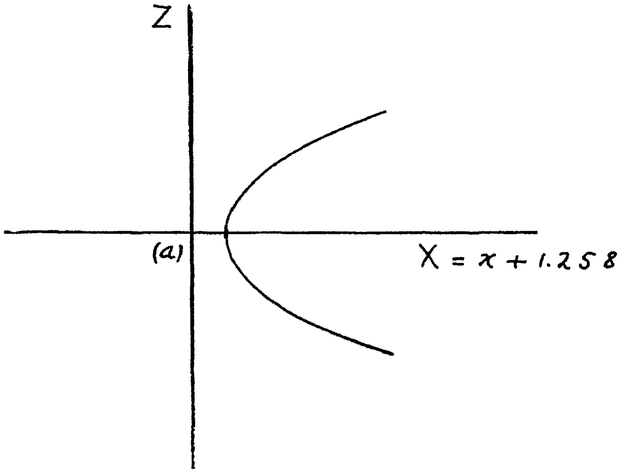


FIGURE 4

yz-plane

$$y = .772z\sqrt{\epsilon^2 - 1.315z^2}.$$

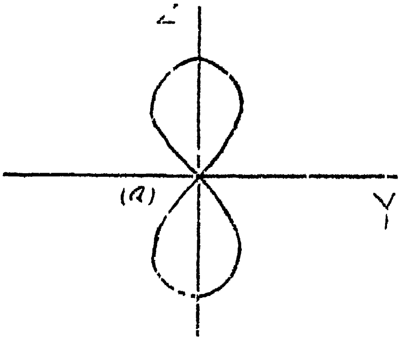


FIGURE 5

The three-dimensional orbit near (*a*) is sketched in Fig. 6.

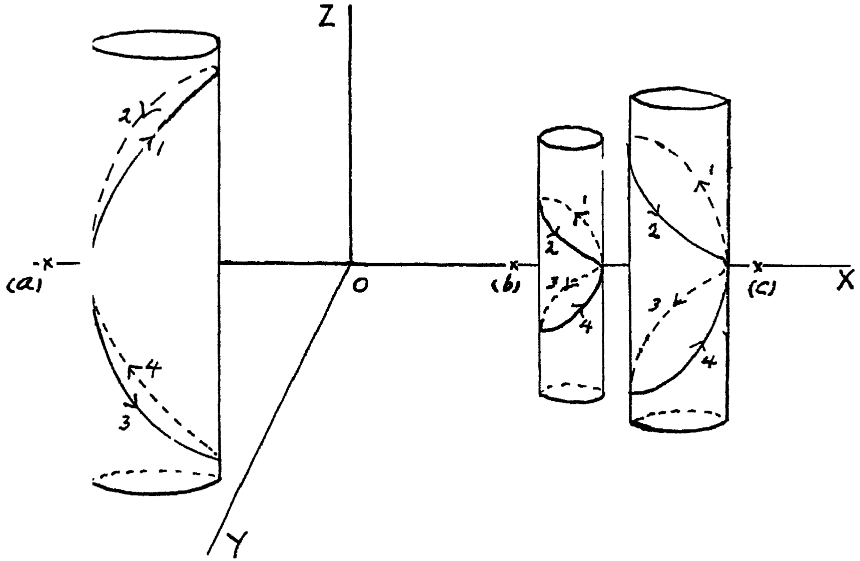


FIGURE 6

Point (*o*)

The orbit in this case is a straight line co-inciding with the z-axis.

Point (*b*)

The equations of the approximate orbit are

$$\begin{aligned} x - .822 &= \epsilon^2 (.992 + .115 \cos 2\gamma\tau), \\ y &= - .046\epsilon^2 \sin 2\gamma\tau, \\ z &= \frac{\epsilon}{\gamma} \sin \gamma\tau. \end{aligned}$$

Point (*c*)

The equations of the approximate orbit are

$$\begin{aligned} x - .916 &= -\epsilon^2 (1.805 - .216 \cos 2\gamma\tau), \\ y &= - .103\epsilon^2 \sin 2\gamma\tau, \\ z &= \frac{\epsilon}{\gamma} \sin \gamma\tau. \end{aligned}$$

The three-dimensional curves near (*b*) and (*c*) are shown in Fig. 6.

For the points (*a*), (*b*), (*c*), the parameter ϵ denotes the initial displacement from the libration points along the x-axis. For the origin, ϵ is proportional to the initial projection along the z-axis.

8. CONVERGENCE OF THE SOLUTIONS

The question of the convergence of the solutions will now be considered.

Periodic solutions may be shown to converge by making use of an existence proof in which dominating series are set up. It was shown by MacMillan (1913), however, that if one can make the formal construction of periodic solutions, similar to those that have been obtained, as power series in a parameter, then such solutions will converge for all values of the time provided the parameter is sufficiently small numerically.

The convergence of asymptotic solutions was considered by Poincaré (1892). He showed that convergence is assured as the time approaches plus infinity provided the solutions have the asymptotic form already defined. Divergence of the solutions would occur should the determinant Δ vanish in the construction of the particular integrals and thus give rise to secular terms in t explicitly. The asymptotic solutions that have been obtained contain no secular terms and consequently they are convergent.

Thus if *periodic* solutions can be *constructed*, their convergence is controlled by a parameter and not by the time; and if *asymptotic* solutions can be *constructed* they will converge as the time becomes infinite.

REFERENCES

BUCHANAN, D. Am. Jour. Math., **39**: 79-110. 1919.
———— Can. Jour. Research, A, **22**: 1-25. 1944.
———— Trans. Camb. Phil. Soc., **22**: 309-340. 1919.
HINRICHSEN, J. J. L. Am. Math. Monthly, **50**: 231-237. 1943.
MACMILLAN, W. D. Trans. Am. Math. Soc., **13**: 146-158. 1913.
MOULTON, F. R. Periodic Orbits, Chap. V, Publication 161. The Carnegie Corporation of America. 1920.
POINCARÉ, H. Les méthodes nouvelles de la mécanique céleste, tome 1, chap. 7. Gautier-Villars, Paris. 1892.

AN ELEMENTARY PROOF OF A THEOREM IN GEOPHYSICS

By T. H. MATTHEWS

Communicated by DAVID A. KEYS, F.R.S.C.

IF the velocity of a seismic wave in the earth is $v_0 + ad$ where v_0 and a are constants and d is the vertical distance below the surface, the quickest wave path between two points A and B on the surface is the arc of a circle in a vertical plane through A and B.

The quickest path lies in the vertical plane through A and B. This can be shown by projecting any other path orthogonally on this plane. For every element ds on the path we have the projection ds' on the plane where $ds' \leq ds$, and the velocity along ds' equals the velocity along ds. Hence the total time along s' is less than the time along s.

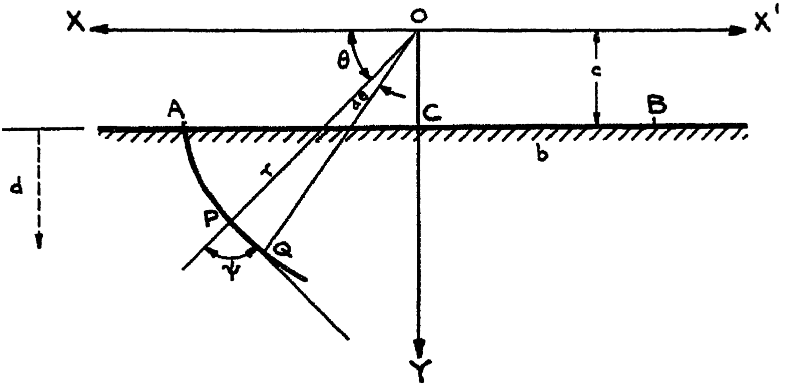


FIGURE 1

Now $v = v_0 + ad = a(d + v_0/a)$.

If we take the axis XX' parallel to AB and a distance $c = v_0/a$ above AB, then $v = ay$.

Let OY bisect AB at C and let $AC = CB = b$.

If P is any point on a wave-front path starting from A, we have

$$\frac{ds}{dt} = v = ay = ar \sin \theta.$$

also $r \dfrac{d\theta}{ds} = \sin \Psi.$

Therefore $\dfrac{d\theta}{dt} = \dfrac{d\theta}{ds} \cdot \dfrac{ds}{dt} = a \sin \theta \sin \Psi \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{I}$

Hence

$\frac{d\theta}{dt}$ is independent of r and is a maximum for a given θ when $\Psi = 90°$

that is, the quickest path from OP to the adjacent line OQ is perpendicular to OP, and the quickest path from OA to OB is a circular arc with centre O and radius OA.

$$\text{Time from } A \text{ to } B = \int dt = \int \frac{d\theta}{a \sin \theta} \text{ from I}$$

$$= \frac{1}{a}\Big[\log \tan \tfrac{1}{2}\,\theta\Big]_{\theta_1}^{\theta_2}$$

$$= \frac{1}{a} \log \frac{\tan \frac{1}{2}\, X\hat{O}B}{\tan \frac{1}{2}\, X\hat{O}A} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\text{II.}$$

Since $\tan X\hat{O}A = \frac{c}{b}$ and $\tan X\hat{O}B = -\frac{c}{b}$ and $c = \frac{v_0}{a}$, we can show

$$\text{that time } = \frac{1}{a} \log \left\{\frac{\sqrt{a^2b^2 + v_0^2} + ab}{\sqrt{a^2b^2 + v_0^2} - ab}\right\} \ldots\ldots\ldots\ \ldots\ \ldots\ldots\text{III.}$$

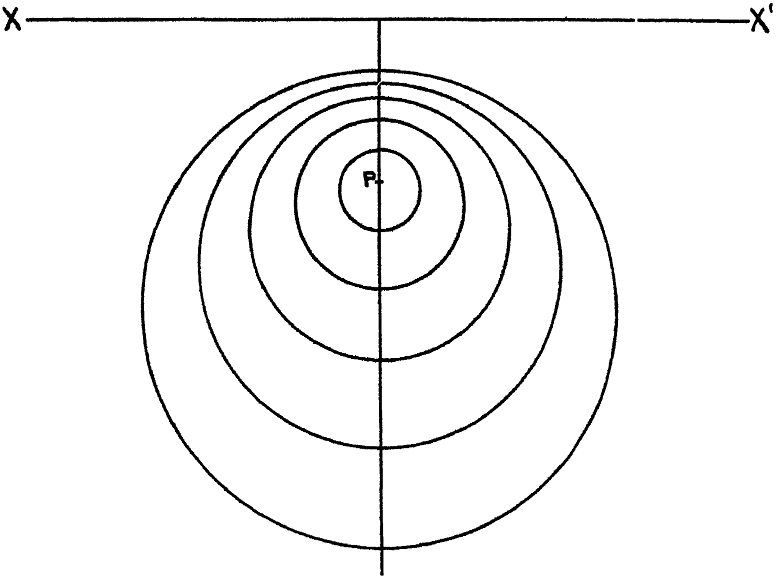


FIGURE 2

Every circle which passes through any point P and has its centre on the fixed line XOX' passes also through P' the image of P in XOX', i.e., these circles form an intersecting coaxal system.

Hence the wave front of a disturbance starting at P is a circle of the orthogonal system which has XOX' as its radical axis.

And, in three dimensions, the wave-front is a sphere belonging to a coaxal system (obtained by rotating the circle about PP').

P is a limiting point (a point sphere) of the system, and as the wave spreads, the centre of the sphere moves vertically downwards from P. Successive stages are shown in Fig. 2.

THE ABSORPTION SPECTRUM OF LIQUID OXYGEN

By H. D. SMITH and N. BARTON

Presented by G. M. SHRUM, F.R.S.C.

(*Read May Meeting, 1942*)

I. *Introduction*

During the course of an attempt to detect in the near infra-red absorption spectrum of liquid hydrogen certain bands predicted by Herzberg (1938) on the basis of the theory of the quadrupole radiation of molecular hydrogen, an apparatus was devised for obtaining long optical path lengths (Smith and Marshall, 1940) in a substance confined in a limited space. Preparatory to working with liquid hydrogen, it was deemed advisable to test the efficiency of the apparatus by employing liquid oxygen, which possesses in the visible and near infrared region a number of well-defined absorption band systems, which have been investigated extensively in recent years (McLennan, Smith, and Wilhelm, 1930; Guillien, 1934, 1936; Ellis and Kneser, 1933). In this test a greater path length of liquid oxygen than had been employed by any previous worker was used. A thorough study was made of the spectral region from λ4200A to λ11500A.

II. *Apparatus*

The general optical arrangement usually employed in obtaining long path lengths in a vertical column of absorbing medium is indicated schematically in Fig. 1. The spherical mirror M_3 is immersed in the medium at the lower end of the container in such a way that two plane mirrors, M_1 and M_2, situated at the other end of the column are at the centre of curvature of M_3. A horizontal beam of light from the source S is condensed by the lens L, and reflected by the small plane mirror M, in a direction designated by the beam A in Fig. 1. Successive reflections of the beam at M_3, M_2, M_1, and again at M_3 give rise to the beams B, C, D as shown in the diagram. The light emerges in a vertical direction between mirrors M_1 and M_2, and after reflection from the small plane mirror M_4, is focussed by a lens on the slit of the spectrograph.

A modification of the above system was used in the investigation of the absorption spectrum of liquid oxygen, and is exhibited in Fig. 2. The spherical mirror M_5 is placed at the lower end of the Dewar flask, with its axis parallel to that of the tube. Two plane mirrors, M_3 and M_4, are suspended from cap *C*, which covers the mouth of the Dewar, and in which there are two circular openings, D_3 and D_4, through which the beam enters and leaves the container. A horizontal beam of light from the carbon arc *S* passes through the diaphragm D_1, and is focussed by lens L_1 on the circular aperture D_2, after being reflected at the plane mirror M_1. The light traverses the length of the flask, and is reflected back along the tube by the mirror M_5. The beam then suffers reflections at M_3 and M_4 which are adjusted so that the beam reflected from M_4 returns to the concave mirror M_5. The light, after its final reflection at M_5, passes through the circular opening D_4 to the mirror M_2 where it is reflected to the lens L_2 which focusses it on the slit of the spectrograph.

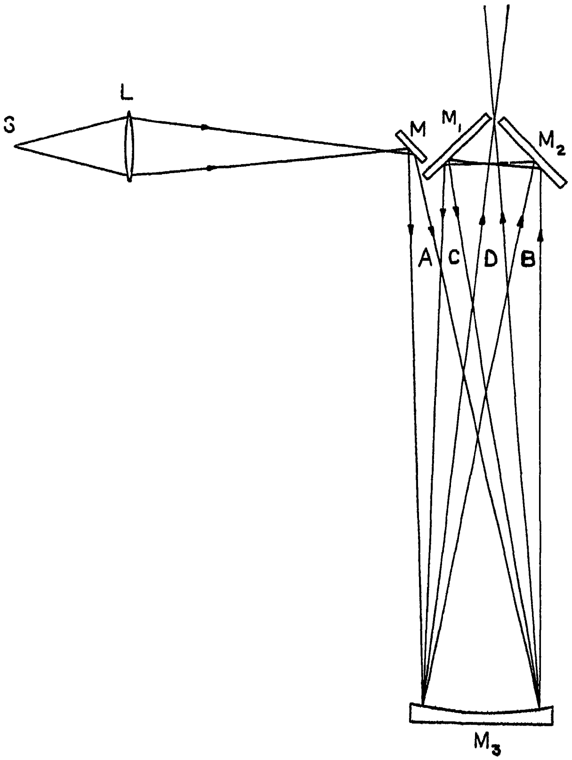


Figure 1

By adjusting mirrors M_3 and M_4 carefully, it is possible to get repeated reflections of the beam at M_3, M_4, and M_5, so that many traversals of the column are obtained. By very careful adjustment a total path length greater than 200 feet has been obtained in a column of air 1 feet long. Owing to the strong absorption of liquid oxygen, the full advantages of the method could not be used in the present experiments. All necessary adjustments of the optical parts within the flask are made from above by means of long bakelite rods. Mirrors used in the apparatus are front-surfaced with evaporated films of aluminum or silver, and mirrors M_3, M_4, M_5 are completely immersed in the absorbing medium.

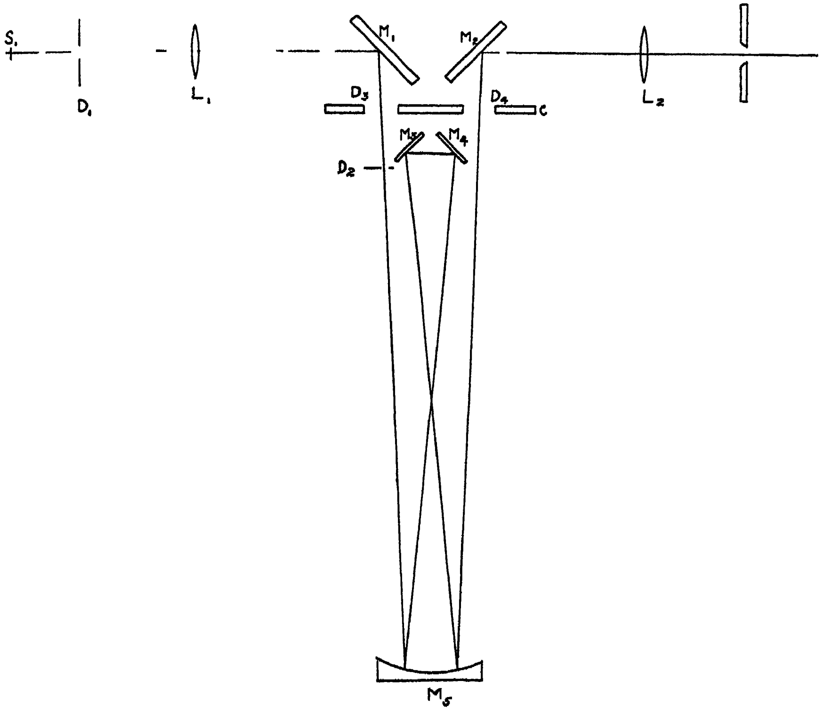


FIGURE 2

One of the chief difficulties encountered in optical experiments involving long columns of liquefied gases at atmospheric pressure, lies in obtaining a quiet surface on the liquid. Even a gentle ebullition gives rise to a disturbance at the surface which diminishes appreciably the amount of light entering or leaving the absorbing medium.

The method devised in this work to keep the surface of the liquid as quiet as possible is exhibited in Fig. 3. The liquid oxygen is contained in a silvered vacuum flask (*A*), 130 cm. long, with an inside diameter of 12 cm.[1] The flask rests on a wooden base *D*, which is lined with a thin layer of felt *C*, the whole being surrounded by a thin-walled sheet metal container *E*, open at the top. A protective layer of felt, which acts also as an absorbent of moisture, is placed between this container and the Dewar. The whole is now placed in a double walled metal container *G*, which is insulated with 3 inches of Kapok. Crushed

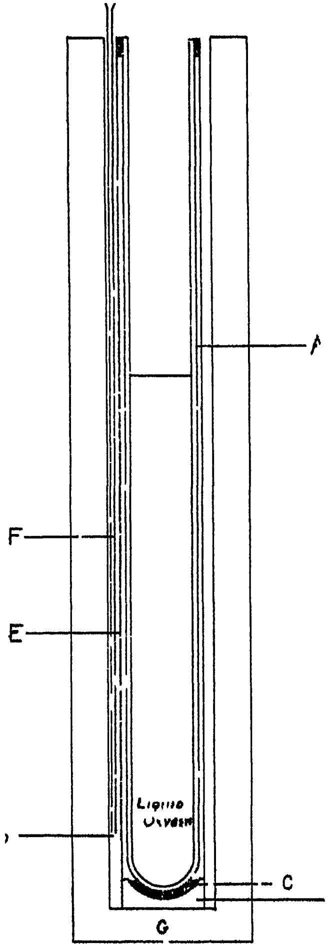


Figure 3

---

[1] This Dewar flask was kindly provided by Professor Giauque, Department of Chemistry, University of California, Berkeley, California.

solid carbon dioxide is placed in the space *B* between the inner and outer metal containers. Additional cooling is provided by pouring liquid air through a small tube *F*, leading to the bottom of the dry ice compartment where it is possible to maintain a small depth of liquid air throughout the experiment. When the system has been cooled by the above means, liquid oxygen may be kept in the vacuum flask for a period of 72 hours or longer.

A Hilger constant deviation spectrograph with a glass prism gave excellent spectrograms, in the region λ4200A. to λ11,500A., on Eastman spectroscopic plates of the types shown in Table I.

TABLE I

Type	*Range*
F	λλ4600 — 6800A.
S	λλ6700 — 7200A.
U*	λλ6800 — 7500A.
R*	λλ7300 — 8500A.
Z*	λλ10000 — 12000A.

*Types U, R, and Z hypersensitized in a solution of 4 per cent ammonia, ethyl alcohol, and water.

III. *Results*

Sixteen absorption bands were obtained in the spectral region investigated in this work. All are degraded towards the violet, and each has a very sharp edge on the long wave-length side. The positions of the band heads are given in the first column of Table II. The results obtained by Guillien (1934, 1936), by Ellis and Kneser (1933), and by McLennan, Smith, and Wilhelm (1930) are tabulated in the second, third, and fourth columns respectively. The three bands recorded in the near infra-red by Ellis and Kneser were obtained by them with an infra-red spectrograph. It is generally accepted as a result of the work of Mulliken (1928), Herzberg (1934), and others that these bands are the first three members of the system arising from transitions between the lower $^3\Sigma$ and the $^1\Delta$ energy levels. The band at λ12,610A. was discovered bolometrically by Abbott in the solar absorption spectrum, and studied further by Herzberg (1934), who established its identity as the $0 \rightarrow 0$ member of the $^3\Sigma$ — $^1\Delta$ system. This band has not been observed photographically. The second member, $1 \rightarrow 0$, at λ10,680A. is not observed in the solar spectrum because of the presence of a strong water vapour absorption band in the same region. As far as the authors are aware, this is the first time that this band has been recorded by photographic means. The next two members of this system, namely, the $2 \rightarrow 0$ at λ9280A., and the $3 \rightarrow 0$ at λ8140A.,

TABLE II

WAVE-LENGTHS OF ABSORPTION BANDS FOR LIQUID OXYGEN

Authors	*Guillien*	*Ellis and Kneser*	*McLennan, Smith, and Wilhelm*
		12610	
10680		10600	
10350			10420
9900			
9280		9190	9200
8140	8150		8250
7640	7630	7650	7665
7060	7307		
6860	6895	6900	6916
6370	6290	6300	6368
5820	5760	5780	5826
5340	5320	5335	5364
4960	4950		4982
4830	4772	4775	4802
4620	4632		4628
4490	4472	4475	4481
4200	4210		4208

appear very clearly on our plates. However, the suggested fifth member of very low intensity at λ7307A, given by Guillien, cannot be detected in our spectrograms even with the greatest path length used. In addition to those of the $^3\Sigma$—$^1\Delta$ system, two other bands were found in the near infra-red region at λ10,350A. and λ9900A. The presence of neither of these can be accounted for satisfactorily on the basis of known energy levels of the O_2 molecule. Due to the difficulty involved in determining accurately the wave-length of the head of this weak band found by the authors at λ10,350A., it is possible that it may be identical with that given by McLennan, Smith, and Wilhelm at λ10,420A.

The present work verifies the existence of all bands previously found in the region from λ7200A. to λ4000A., and in addition furnishes evidence of a very weak band at λ7060A.

Guillien did not observe visually on his plates the weak band at λ4826A. first given by McLennan, Smith, and Wilhelm, but thought that his microphotometer tracings indicated the possibility of a very weak band at λ4632A. The present work, however, definitely verifies the existence of this band which can be observed visually and which has a wave-length of λ4620A. on its long wave-length side.

It is interesting to compare spectrograms obtained with different path lengths of liquid oxygen. As we go to greater amounts of absorbing material very noticeable increases occur in the widths of some of the principal bands. This is illustrated in Plate I, where (*a*), (*b*), (*c*) are spectrograms of the visible region for thicknesses of 32 cm., 130 cm., and 180 cm., respectively, of liquid oxygen. In particular, the bands with heads at λ5820A. and λ6370A. increase gradually in width until they each extend over 350A. In fact, for a thickness of 180 cm. of liquid oxygen there is almost complete absorption in the spectral region from λ6365A. to λ5200A.

The experiments described above were carried out at the University of California at Berkeley, California, and the authors are indebted to the Physics and Chemistry Departments of that institution for the use of their facilities. Thanks are due particularly to Professors Birge, Giauque, Jenkins, and White for their kindness in providing special equipment.

Physics Department,
University of British Columbia,
Vancouver, B.C.
May 7, 1945.

REFERENCES

Ellis, J. W. and H. O. Kneser. Combinational Relations in the Absorption Spectrum of Liquid Oxygen. Zeits f. Physik, **86**: 583-86. 1933.

Guillien, R. Absorption of Long Columns of Liquid Oxygen. Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences, Paris, **198**: 1223-25. 1934.

———— Existence of O_4 in Liquid Oxygen. Comptes Rendus, **198**: 1486-88. 1934.

———— Absorption Bands of Liquid Oxygen. Comptes Rendus, **202**: 1373-75. 1936.

———— The 7635 A. Band of Liquid Oxygen. Comptes Rendus, **202**: 1423-25. 1936.

Herzberg, G. Photography of the Solar Spectrum to Wave Length 12,900 A. Nature, **133**: 759. 1934.

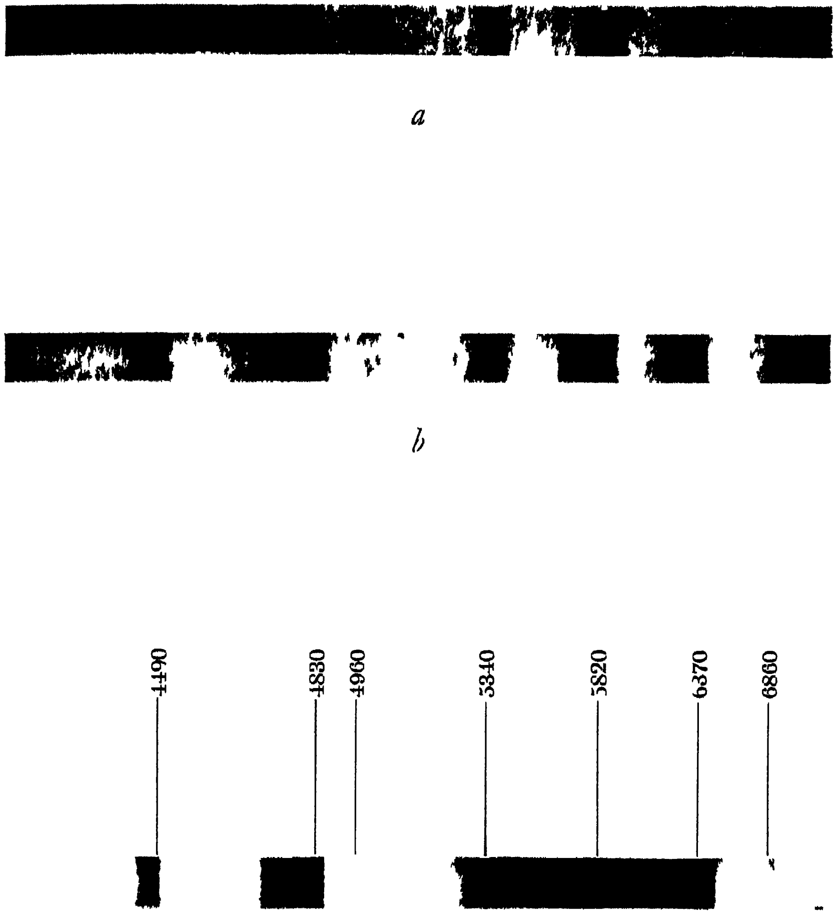
———— Possibility of Detecting H_2 and N_2 in Planetary and Stellar Atmospheres by their Rotational-Vibration Spectra. Astrophysical Journal, **87**: 428-37. 1938.

McLennan, J. C., H. D. Smith & J. O. Wilhelm. On the Energy Levels of Molecular Oxygen. Trans. Roy. Soc. Can., Sec. III, **24**: 65-89. 1930.

Mulliken, R. S. Interpretation of the Atmospheric Absorption Bands of Oxygen. Phys. Rev., **32**: 880-87. 1928.

Smith, H. D. & J. K. Marshall. Method for Obtaining Long Optical Paths. J.O.S.A., **30**: 338-342. 1940.

PLATE I



a

b

c

Absorption spectrum of liquid oxygen

Length of path (*a*) 32 cm, (*b*) 130 cm (*c*) 180 cm

WAVE-IMPEDANCES AND THE EFFECTIVE CROSS-SECTIONS OF ANTENNAS

By W. H. WATSON, F.R.S.C.

SUMMARY

The paper connects wave conceptions, circuit ideas, and notions arising in the theory of the scattering of particles. Wave-impedance, a more general concept than field-impedance, is defined (§1) The connection of wave-impedances with the effective cross-section of a scattering antenna is developed for waves defined by a scalar wave-function (§2), and for electromagnetic waves (§3). Matrix representation of coupling between different types of waves is discussed (§4). The method is extended to include the treatment of an antenna connected to a receiver or transmitter (§5). The last section discusses high gain antennas by a method analogous to that used by Heisenberg (1943) for atomic particles.

By suppressing the full geometrical implication of field theory, the aim of this work is to extract in antenna-wave problems the system of connections analogous to circuit theory.

§ 1. *Wave-Impedance*

The ideas to be presented in this paper form a link between wave conceptions, circuit ideas, and notions which are most naturally developed in connection with the scattering of streams of particles. The aim is to extract the system of connections in antenna-wave problems by suppressing the full geometrical implications of field theory in the same kind of way that is followed in treating lumped loads on transmission lines.

In the reflexion of the principal wave on a linear parallel-plane strip transmission line, the field impedance, defined as the ratio electric force/magnetic force, varies with position along the line. At the termination it is equal to the impedance presented to the line when the strips are of unit width and are unit distance apart. In this simple type of wave, the electric and magnetic forces do not vary in planes perpendicular to the direction of propagation. In a coaxial line they do vary, but the form of their dependence on distance from the axis is exactly the same for both, and in the principal wave the distribution possesses circular symmetry about the axis. Hence the ratio E_r/H_θ is independent of position in a cross-section of the line between the two conductors. This principle may be applied to any characteristic wave; namely, if E and H are at right angles to each other, and for the particular wave considered, depend on the coordinates specifying position on a surface of equal phase by exactly the same function of these coordinates, the conception field-impedance applied to the particular characteristic wave in question is perfectly clear, and as Schelkunoff (1937, 1938) has shown, can be applied usefully to the plane waves in wave-guides and to the spherical and cylindrical waves in space. With plane waves, the characteristic field-impedance does not depend on distance in the direction of propagation, whereas with the spherical and cylindrical waves it does.

The fact that impedance is defined as E/H is in large measure irrelevant so far as waves are concerned, for essentially the same properties apply in determining the reflexion, etc., of any waves of one characteristic form. In fact there is a definite advantage in concentrating attention on the conception wave-impedance which has simpler properties than field-impedance and will now be defined.

Suppose that for simplicity we consider a plane wave satisfying

$$\nabla^2 u = \frac{1}{c^2}\frac{\partial^2 u}{\partial t^2}, \tag{1}$$

the propagation being in the z-direction. For the harmonic characteristic distribution

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = -\chi^2 u, \quad \frac{\partial^2 u}{\partial t^2} = -\omega^2 u \tag{2}$$

we have

$$\frac{\partial^2 u}{\partial z^2} = \left(\chi^2 - \frac{\omega^2}{c^2}\right)u = -k^2 u; \tag{3}$$

and, suppressing the time factor $e^{j\omega t}$

$$u = Ae^{-jkz} + Be^{+jkz}. \tag{4}$$

The first term represents a wave travelling in the direction z-increasing, the second a wave travelling in the opposite direction, while the complex numbers A and B yield the amplitudes and phases of these waves at $\omega t = kz$ and $\omega t = -kz$ respectively.

Suppose that this wave system exists for $z \leq 0$, $z = 0$ being the termination of the "line" where the waves are reflected, then B/A denotes the reflexion coefficient and $\dfrac{A+B}{A-B}$ is the impedance, both being reckoned at $z = 0 \left(\mathrm{mod}\,\dfrac{\lambda}{2}\right)$ where λ is the wave-length on the line. The reader will easily verify for himself that this result is quite in accordance with circuit ideas applied to a conventional transmission line, provided that he multiplies $\dfrac{A+B}{A-B}$ by the characteristic impedance (Z_0) of the line. Now the way in which Z_0 enters the wave picture is through the idea of energy flow.

Consider equation (3). Introduce the function v so that

$$-\frac{\partial u}{\partial z} = pv \qquad -\frac{\partial v}{\partial z} = qu. \tag{5}$$

Then

$$pq = -k^2. \tag{6}$$

In a pure travelling wave in the direction z-increasing

$$u = Ae^{-jkz}, \quad v = Aj\frac{k}{p}e^{-jkz} \tag{7}$$

and

$$\frac{v}{u} = j\frac{k}{p} = \sqrt{\frac{q}{p}}\,. \tag{8}$$

If, in this wave, the mean energy flux density $\overline{W}$ in the direction of propagation is $\frac{1}{2}\,u^k v$ where u^k is the conjugate complex to u

$$\overline{W} = \tfrac{1}{2}A^2\sqrt{\frac{q}{p}}\,. \tag{9}$$

$\sqrt{\frac{q}{p}}$ is the characteristic field-impedance of the wave in question. If v represents electric force, u magnetic force, then this quantity is measured in ohms. From the physical point of view, however, the meaning of characteristic impedance is relative to the physical character of the wave-function the square of whose amplitude determines the energy flux.

So long as it is understood that we refer wave-impedance to the appropriate characteristic impedance for the wave considered, we need concern ourselves only with the dimensionless ratio $\frac{A+B}{A-B}$ that is, by omitting from field-impedance the number (or function of the space-coordinate in the direction of propagation) characteristic of the wave in question, we arrive at wave-impedance which depends only on the relation, as to both amplitude and phase, of the waves travelling in oposite directions. For a spherical wave, for example, with the wave-function satisfying (1), the radial functions are[1]

$$u_n^+ = \sqrt{\frac{\pi}{2kr}}\,H_{n+\frac{1}{2}}^{(1)}(kr) \text{ and } u_n^- = \sqrt{\frac{\pi}{2kr}}\,H_{n+\frac{1}{2}}^{(2)}(kr)$$

and in place of (4) we write

$$u_n = A_n u_n^+ + B_n u_n^-. \tag{10}$$

---

[1] See §2 for the definition of these functions.

u_n^+ is the outgoing and u_n^- the incoming wave of the nth order. The ratio $(A_n + B_n)/(A_n - B_n)$ is then the wave-impedance in this nth order wave system, viewed outwards from the centre of the sphere. This impedance is independent of distance from the origin. The reader is referred to Schelkunoff's paper "The Transmission Theory of Spherical Waves" where he will observe that field-impedance is a function of distance from the origin and that we are therefore required to think of a non-uniform transmission line.

The term wave-admittance easily finds expression as $(A - B)/(A + B)$; it is understood that the characteristic wave in question has its own characteristic admittance, which is the field admittance in a pure travelling wave. It will generally be a function of the coordinate used to specify propagation of phase.

§ 2. *Effective Cross-Sections*

As a step towards the treatment of electromagnetic waves in space, we consider waves specified by means of a scalar wave-function, in order to avoid the additional complication of the vector wave-functions required to represent electromagnetic waves. Imagine a uniform plane wave propagated parallel to z and incident on an "antenna" situated at the origin. This antenna is excited by the incident wave and by radiating scatters it, so that energy is removed from the incident directed beam.

Let the incident harmonic wave (time factor $e^{-j\omega t}$) be represented by $ae^{jkr\cos\theta}$ where θ is measured from the polar axis coincident with the direction of propagation. This wave may be represented as the superposition of waves converging on and diverging from the origin as follows:

$$ae^{jkr\cos\theta} = a\sum_{n=0}^{\infty} j^n \frac{2n+1}{2}\sqrt{\frac{\pi}{2kr}} \left(H^{(1)}_{n+\frac{1}{2}}(kr) + H^{(2)}_{n+\frac{1}{2}}(kr)\right) P_n(\cos\theta) \tag{11}$$

where P_n denotes the Legendre polynomial of order n, and the H-functions are the Hankel functions which have the following asymptotic forms $(r \to \infty)$

$$\sqrt{\frac{\pi}{2kr}}\, H^{(1)}_{n+\frac{1}{2}}(kr) = \frac{1}{kr}\, e^{j\left(kr - \frac{n+1}{2}\pi\right)}$$

$$\sqrt{\frac{\pi}{2kr}}\, H^{(2)}_{n+\frac{1}{2}}(kr) = \frac{1}{kr}\, e^{-j\left(kr - \frac{n+1}{2}\pi\right)}. \tag{12}$$

When one considers the time factor $e^{-j\omega t}$, it is evident that these functions represent respectively at great distances from the origin outgoing and incoming spherical waves.

We shall write

$$j^n \sqrt{\frac{\pi}{2kr}}\, H^{(1)}_{n+\frac{1}{2}}(kr) P_n(\cos\theta) = O_n$$

$$j^n \sqrt{\frac{\pi}{2kr}}\, H^{(2)}_{n+\frac{1}{2}}(kr) P_n(\cos\theta) = I_n. \tag{13}$$

Thus equation (11) becomes

$$P = ae^{jkr\cos\theta} = a \sum_{n=0}^{\infty} \frac{2n+1}{2}(O_n + I_n). \tag{11'}$$

On the assumption that only zonal spherical waves are radiated, the outgoing scattered wave is similarly represented as

$$S = \sum_{n=0}^{\infty} c_n O_n \tag{14}$$

where the coefficients c_n are as yet undetermined.

The wave system resulting from the superposition of P and S is

$$P + S = \sum_{n=0}^{\infty} \left[\left(c_n + \frac{2n+1}{2} a \right) O_n + a \frac{2n+1}{2} I_n \right] \tag{15}$$

$$= \sum_{n=0}^{\infty} (A_n I_n + B_n O_n). \tag{16}$$

We define $B_n/A_n = w_n$ as the reflexion coefficient (looking toward the antenna) for the nth characteristic wave chosen. It is necessary, of course, that (15) shall represent, outside a sufficiently large sphere, the solution of the wave-equation subject to the boundary conditions imposed by the presence of the antenna in the incident plane wave. The representation of the solution in the form (15) is unique as is well known from the theory of orthogonal functions.

Since $$B_n = \frac{2n+1}{2} a + c_n \text{ and } A_n = \frac{2n+1}{2} a \tag{17}$$

$$c_n = \frac{2n+1}{2} a(w_n - 1). \tag{18}$$

The fraction of incident energy flux density scattered by the antenna through a distant sphere is

$$\sigma = \frac{1}{a^2}\iint SS' r^2 \sin\theta \, d\theta \, d\varphi = \frac{1}{a^2}\sum_{n=0}^{\infty} \frac{c_n c_n'}{k^2} \cdot \frac{4\pi}{2n+1} \cdot \tag{19}$$

Thus

$$\sigma = \sum_{n=0}^{\infty} \frac{\pi}{k^2} (2n+1) \, | w_n - 1 |^2 \text{ metres}^2 \tag{20}$$

is the area which is called the effective cross-section of the antenna. If presented to a uniform directed beam of particles containing $1/\sigma$ particles per unit area this target would remove one particle from the beam. The idea is well known in atomic physics.

(i) Suppose that $w_n = e^{2j\delta_n}$ corresponding to the purely reactive wave-impedance

$$\frac{1 + e^{2j\delta_n}}{1 - e^{2j\delta_n}} = -j \cot \delta_n.$$

Then

$$|w_n - 1|^2 = 4 \sin^2 \delta_n$$

and

$$\sigma = \sum_{n=0}^{\infty} \frac{4\pi}{k^2} (2n+1) \sin^2 \delta_n; \tag{21}$$

σ is then a measure of the effective cross-section presented to the incident wave when no energy is absorbed by the scattering antenna, for the wave-impedance has no real part.

(ii) More generally let $|w_n| < 1$ and let

$$\frac{1 + w_n}{1 - w_n} = z_n = r_n + j x_n.$$

Then

$$| w_n - 1 |^2 = \left| \frac{2}{z_n + 1} \right|^2 . \tag{22}$$

The value of σ given by equations (20) and (22) now represents the effective cross-section for reradiation. The cross-section for the total energy absorbed from the incident beam is obtained by substituting $1 - |w_n|^2$ for $|1 - w_n|^2$ in equation (20). Thus the total cross-section in this case is

$$\sigma = \sum_{n=0}^{\infty} \frac{4\pi}{k^2} (2n+1) \frac{r_n}{(r_n + 1)^2 + x_n^2} . \tag{23}$$

§ 3. *Electromagnetic Waves*

In the case of electromagnetic waves we consider a plane linearly polarized wave propagated in the z-direction, and incident on an antenna. All the cartesian components of the field of the incident wave vanish except

$$E_x = ae^{jkz} \quad H_y = \frac{a}{Z_0} e^{jkz}, \qquad \left(Z_0 = \sqrt{\frac{\mu_0}{\kappa_0}}\right) \tag{24}$$

(the time-factor $e^{-j\omega t}$ being omitted as usual).

In spherical polar coordinates (r, θ, ϕ) referred to z as the polar axis, centred on some suitable point in the antenna and with $\varphi = 0$ on the xz-plane, we obtain for the radial components

$$\begin{aligned} E_r &= ae^{jkr\cos\theta}\sin\theta\cos\phi \\ H_r &= \frac{a}{Z_0} e^{jkr\cos\theta}\sin\theta\sin\phi \end{aligned} \tag{25}$$

The expressions (25) can be developed in the form

$$\begin{aligned} e^{jkr\cos\theta}\sin\theta \begin{smallmatrix}\cos\phi\\ \sin\phi\end{smallmatrix} = \sum_{n=1}^{\infty} \frac{(2n+1)j^{(n-1)}}{2kr}\sqrt{\frac{\pi}{2kr}}\left[H^{(1)}_{n+\frac{1}{2}}(kr) + H^{(2)}_{n+\frac{1}{2}}(kr)\right] \\ \times P^{(1)}_n(\cos\theta)\begin{smallmatrix}\cos\phi\\ \sin\phi\end{smallmatrix} \end{aligned} \tag{26}$$

analogous to (11) above, tesseral harmonics replacing zonal harmonics. Each of the terms of this series has the form of a characteristic distribution of the radial electric or magnetic force in transverse magnetic and transverse electric spherical waves in free space. The other field components are given by the following formulae (Frenkel, 1926; Stratton, 1941):

Transverse Magnetic Waves (nth order) $\quad H_r = 0$

$$\begin{aligned} E_\theta &= \frac{1}{n(n+1)}\,\frac{1}{r}\,\frac{\partial^2}{\partial\theta\partial r}(r^2E_r) & H_\theta &= \frac{j\omega\kappa_0}{n(n+1)}\,\frac{1}{r\sin\theta}\,\frac{\partial}{\partial\phi}(r^2E_r) \\ E_\phi &= \frac{1}{n(n+1)}\,\frac{1}{r\sin\theta}\,\frac{\partial^2}{\partial\phi\partial r}(r^2E_r) & H_\phi &= \frac{-j\omega\kappa_0}{n(n+1)}\,\frac{1}{r}\,\frac{\partial}{\partial\theta}(r^2E_r). \end{aligned} \tag{27a}$$

Transverse Electric Waves (nth order) $\quad E_r = 0$

$$\begin{aligned} H_\theta &= \frac{1}{n(n+1)}\,\frac{1}{r}\,\frac{\partial^2}{\partial\theta\partial r}(r^2H_r) & E_\theta &= \frac{-j\omega\mu_0}{n(n+1)}\,\frac{1}{r\sin\theta}\,\frac{\partial}{\partial\phi}(r^2H_r) \\ H_\phi &= \frac{1}{n(n+1)}\,\frac{1}{r\sin\partial}\,\frac{\partial^2}{\partial\phi\partial r}(r^2H_r) & E_\phi &= \frac{j\omega\mu_0}{n(n+1)}\,\frac{1}{r}\,\frac{\partial}{\partial\theta}(r^2H_r). \end{aligned} \tag{27b}$$

Now outside a sphere of sufficiently great radius—this is determined by the size of the antenna—the scattered wave radiated by the antenna can be represented by

$$S = E_s = \sum_{n,m} \frac{c_{nm}}{(kr)^{3/2}} H^{(1)}_{n+\frac{1}{2}}(kr)\, j^{n-m} T_{nm}(\theta, \phi) \quad \text{(T.M. type)} \quad (28a)$$

$$S' = H_s = \sum_{n,m} \frac{c'_{nm}}{(kr)^{3/2}} H^{(1)}_{n+\frac{1}{2}}(kr)\, j^{n-m} T'_{nm}(\theta, \phi) \quad \text{(T.E. type)} \quad (28b)$$

where T_{nm} and T'_{nm} are tesseral harmonics.

For the purpose of exposition, the explicit notation which we have so far employed is much too cumbersome; it is convenient to use a single letter s to distinguish the characteristic waves. Imagine the wave-functions arranged in the following order: the tesseral harmonics of each group with the same value of n are placed together in the order $\cos n\phi, \cos (n-1)\phi, \ldots\ldots \cos \phi, 1, \sin \phi, \sin 2\phi, \ldots\ldots \sin n\phi$. We shall conceive of the groups as arranged according to increasing n. Further we shall think of the series of electric wave-functions preceding the magnetic ones. Thus a wave-function of electric type (T.M.-wave) is denoted by E_s^+ for an outgoing wave and E_s^- for the incoming wave with the same angular distribution; for a magnetic wave-function we shall use H_s^+ and H_s^- respectively. When we wish to denote a wave-function belonging to the complete set of electric and magnetic type wave-functions, we shall use the symbol F_s. When we wish to exclude from the range of s those wave-functions for which $m = 1$ we shall use σ in place of s, and to denote waves in which $m = 1$ we shall use the Greek letter ν which is to take the appropriate value of n. We shall further understand that all the E_s and H_s wave-functions are normalized in the following way.

For the outgoing T.M.-spherical wave in which the radial electric force E_r has the asymptotic form

$$\frac{A}{(kr)^2} e^{jkr} T_{nm}(\theta, \phi), \qquad (29)$$

the power radiated is on the average

$$\frac{1}{2} \frac{|A|^2}{k^2 Z_0 n(n+1)} \cdot \frac{4\pi}{2n+1} \cdot \frac{(n+m)!}{(n-m)!} = \frac{1}{2} \frac{|A|^2}{Z_0 k^2} p_{nm}. \qquad (30)$$

For the outgoing T.E.-spherical wave in which $Z_0 H_r$ has the asymptotic form (29), the average power radiated is also given by (30). We note that

$$p_{nm} = \frac{4\pi}{n(n+1)(2n+1)} \frac{(n+m)!}{(n-m)!}, \quad p_{n1} = \frac{4\pi}{2n+1}. \qquad (31)$$

We normalize by dividing the functions previously used in equations (26) and (28) by $\sqrt{p_{nm}}$ for electric type and by $Z_0\sqrt{p_{nm}}$ for magnetic type waves. Thus equations (26) being substituted in (25)

$$P = \sum_{\nu=1}^{\infty} a_\nu (E_\nu^+ + E_\nu^-) \tag{32a}$$

$$P' = \sum_{\nu=1}^{\infty} a_\nu (H_\nu^+ + H_\nu^-) \tag{32b}$$

where

$$a_\nu = \frac{a(2n+1)}{2}\sqrt{p_\nu}, \tag{33}$$

and it is understood that the E_ν depend on ϕ through cos ϕ, the H_ν through sin ϕ. P denotes the incident radial electric force, P' the incident radial magnetic force. We shall combine (32a) and (32b) in the form

$$P = \sum_{\nu=1}^{\infty}{}' a_\nu (F_\nu^+ + F_\nu^-), \tag{34}$$

where the prime denotes summation over both electric and magnetic types. It is to be clearly understood that the contracted form (34) should be read as a vector equation in which the electric and magnetic type expressions do not add numerically to give P; the ordinary numerical sum of electric and magnetic terms is meaningless for our purpose.

The scattered wave we now write as

$$S = \sum_{\sigma}{}' c_\sigma F_\sigma^+ + \sum_{\nu=1}^{\infty}{}' c_\nu F_\nu^+]. \tag{35}$$

The wave resulting from the superposition of (34) and (35) is

$$S + P = \sum_{\sigma}{}' c_\sigma F_\sigma^+ + \sum_{\nu=1}^{\infty}{}' [(a_\nu + c_\nu) F_\nu^+ + a_\nu F_\nu^-]. \tag{36}$$

In order to express the c_σ in terms of the a_ν we must know the coupling coefficients which represent the action of the antenna in converting incoming waves into outgoing waves of differing mode. When the antenna is excited by the incoming system $\sum_s a_s F_s^-$ let the outgoing system be

$$\sum_{r}{}' \sum_{s}{}' \gamma_{rs} a_s F_r^+. \tag{37}$$

The complex numbers γ_{rs} are the coupling coefficients: the form of (37) is justified by the linearity of electromagnetic laws. From (34) and (37) we deduce

$$c_\sigma = \sum_{\nu=1}^{\infty}{}' \gamma_{\sigma\nu} a_\nu \tag{38}$$

while, using w_ν to denote the complex reflexion coefficients,

$$c_\nu = a_\nu (w_\nu - 1). \tag{39}$$

The total power scattered in both types of waves is

$$\frac{1}{2Z_0k^2}\sum_s{}' |c_s|^2 \tag{40}$$

$$= \frac{a^2}{2Z_0k^2}\sum_\nu{}' \pi(2\nu+1)\left\{\sum_\sigma{}'|\gamma_{\sigma\nu}|^2+|w_\nu-1|^2\right\}. \tag{41}$$

Hence the effective cross-section of the antenna is

$$\sigma = \frac{\pi}{k^2}\sum_\nu{}'(2\nu+1)\left\{\sum_\sigma{}'|\gamma_{\sigma\nu}|^2+|w_\nu-1|^2\right\}. \tag{42}$$

If no energy is lost in the antenna itself, the energy taken from the νth electric mode, for example, in the incident wave being all re-radiated, we have

$$\sum_\sigma{}'|\gamma_{\sigma\nu}|^2+|w_\nu-1|^2 = 1-|w_\nu|^2.$$

So
$$1+\sum_\sigma{}'\frac{|\gamma_{\sigma\nu}|^2}{|w_\nu-1|^2} = \frac{1-|w_\nu|^2}{|1-w_\nu|^2} = r_\nu \tag{43}$$

where r_ν is the resistance or real part of the wave-impedance presented to the νth incident spherical wave of one type. Hence

$$\sigma = \frac{4\pi}{k^2}\sum_\nu{}'(2\nu+1)\frac{r_\nu}{(r_\nu+1)^2+x_\nu^2}, \tag{44}$$

which is entirely analogous to (23).

§ 4. *Coupling Coefficients and Matrix Representation*

From equation (39) we have

$$\gamma_{\nu\nu} = (w_\nu - 1), \tag{45}$$

which immediately suggests that we should redefine the coupling coefficients entering the problem we are considering as follows.

Let
$$g_{s\nu} = \frac{\gamma_{s\nu}}{\gamma_{\nu\nu}}. \tag{46}$$

We then have the relations

$$c_\sigma = \sum_\nu{}' g_{\sigma\nu}c_\nu. \tag{47}$$

We shall use $C(s)$ to denote the column matrix of the c's arranged in the order of the wave-functions. Then

$$C(\nu) = DC(s), \tag{48}$$

where D is the diagonal matrix with zero elements except for those values of s corresponding to $m = 1$. It is a matrix which represents the incident wave.

Since there is no singularity in this wave at the antenna, for it is conceived in the absence of the latter, both outgoing and incoming waves of each mode must be represented in it and be of equal amplitude.

We now write (47) in matrix form

$$C(s) = GC(\nu) = GDC(s). \tag{49}$$

G is the matrix whose elements are g_{sr}. It is easy to be tempted to imagine that the matrix G summarizes the properties of the antenna, but this is not so. Nor is it true of the matrix Γ whose elements are γ_{rs}. G is dependent on the mode of oscillation of the antenna. Hence the system of coupling coefficients is characteristic of a particular mode of oscillation of the antenna. If we are dealing with a resonant structure, the mode of oscillation at or near resonance is the dominant one and in this case the coupling matrix G could be obtained, except for a constant factor, by analysis of the radiation pattern from the antenna when it is excited at its driving point.

In terms of the g's, equation (42) becomes

$$\sigma = \frac{\pi}{k^2} \Sigma'(2\nu+1)\,|w_\nu - 1|^2 \left\{1 + \sum_{\sigma}{}' \,|g_{\sigma\nu}|^2\right\} \tag{50}$$

and (43), for the case of the antenna in which all energy absorbed is reradiated, becomes

$$1 + \sum_{\sigma}{}' \,|g_{\sigma\nu}|^2 = r_\nu. \tag{51}$$

These considerations bring us to the question as to how we represent an antenna coupled by its feed to some other system.

§ 5. *Receiving or Transmitting Antenna*

In the feed, only a single wave is present and can be propagated: the higher order modes of propagation in the feed-line do not enter in the transfer of energy along lengths of line of practical importance. Consequently the effects of the possibility of attenuated propagation in these higher modes can be represented as reactive lumped loading of the line regarded as carrying a single wave. When a generator is delivering energy to the antenna which in turn radiates it, the antenna presents a certain load to the line and in general reflects part of the wave; we shall use w_0 to denote the complex reflexion coefficient on the feed-line, viewed from the driving point.

Now so long as we are concerned with representations like (32) in which addition of waves stands for their superposition in space, we evidently cannot incorporate waves on the feed-line in our system.

But when we employ representations in the form (34) with the corresponding matrices, we see that F_s^+, F_s^- can be regarded as basic vectors in the type of space used in quantum-mechanical representation, one dimension for each characteristic outgoing and incoming spherical wave. Any wave in the physical space of three dimensions is represented by a column matrix or vector in the ∞-dimensional space. Clearly we may adjoin to the system of basic vectors the pair of normalized wave-functions for the feed-line and their complex amplitudes. Then we extend the matrices Γ and G to include the coupling of the waves on the feed-line to the characteristic spherical waves.

The results we have obtained can now be modified to treat the cases when the antenna is connected to a receiver of given input impedance—although one usually expects the receiver to be matched to the feed-line—and also when the antenna is connected to a generator specified in the usual circuit manner by means of its e.m.f. and impedance. This latter case includes the former when the e.m.f. vanishes.

In order to illustrate the application of these principles let us consider a centre-fed half-wave wire antenna, whose axis points in the direction $\theta = \alpha$, $\phi = \beta$, the frequency of the incident plane wave being near to resonance in the antenna's lowest order mode. In this mode the antenna, when driven in the absence of the plane wave, radiates approximately an electric type wave with angular dependence $P_1(\cos\theta')$ where θ' is measured from the polar axis of the antenna. Since

$$\cos\theta' = \cos\alpha\cos\theta + \sin\alpha\sin\theta\cos(\phi-\beta)$$

$$\begin{aligned} P_1(\cos\theta') = \cos\alpha\, P_1(\cos\theta) &+ \sin\alpha\sin\beta P_1^{(1)}(\cos\theta)\cos\phi \\ &- \sin\alpha\sin\beta P_1^{(1)}(\cos\theta)\sin\phi. \end{aligned} \tag{52}$$

There are four waves involved, therefore, in the oscillation of the antenna and we shall refer to them by the scheme

1	2	3	0
$P_1^{(1)}(\cos\theta)\cos\phi$	$P_1^{(1)}(\cos\theta)\sin\phi$	$P_1(\cos\theta)$	feed.

The coupling coefficients which we require are

$$\gamma_{01},\ \gamma_{11},\ \gamma_{21},\ \gamma_{31},\ \gamma_{10},\ \gamma_{20},\ \gamma_{30},\ \gamma_{00}.$$

Since the waves 1, 2, 3 are always radiated in the proportion

$$\sin\alpha\cos\beta : -\sin\alpha\sin\beta : \cos\alpha.$$

We write

$$\left.\begin{aligned} \gamma_{10} &= r\sin\alpha\cos\beta & \gamma_{11} &= q\sin\alpha\cos\beta \\ \gamma_{20} &= -r\sin\alpha\sin\beta & \gamma_{21} &= -q\sin\alpha\sin\beta \\ \gamma_{30} &= r\cos\alpha & \gamma_{31} &= q\cos\alpha \end{aligned}\right\} \tag{53}$$

and we have
$$\begin{aligned} c_1 &= \sin\alpha\cos\beta(qa_1+ra_0) \\ c_2 &= -\sin\alpha\sin\beta(qa_1+ra_0) \\ c_3 &= \cos\alpha(qa_1+ra_0) \\ c_0 &= \gamma_{01}a_1+\gamma_{00}a_0 \end{aligned} \tag{54}$$

where $\gamma_{00}=\dfrac{z_A-z_L}{z_A+z_L}$, z_A being the impedance presented to the feed-line, and z_L the characteristic impedance of the line.
When the antenna is connected to a matched receiver
$$|c_0|^2=|\gamma_{01}|^2|a_1|^2 \tag{55}$$
so that $|\gamma_{01}|$ can be measured by determining the energy picked up from a plane wave of known intensity. The modulus of the coefficient r is readily obtained by feeding the antenna, for the energy radiated is proportional to
$$\sum_{s=1,2,3}|c_s|^2=|r|^2a_0'. \tag{56}$$
If the antenna is excited by a generator of e.m.f. V_0 and impedance z_g (referred to the characteristic impedance of the feed-line as unity) we have, using primes to denote the transformation of a_0 and c_0 due to displacement from the antenna feed-point to the other end of the feed-line,
$$a_0'+c_0'=V_0-z_g(a_0'-c_0'). \tag{57}$$

This equation yields a_0 in terms of V_0 and the amplitude of the incident plane wave. Thus provided we know q, r, γ_{01} and z_A, the radiation, etc., is determined.

The coupling coefficients being complex numbers each require two measurements to determine them. Nevertheless it is not beyond the bounds of possibility that such measurements could actually be made with models operated at microwave frequencies, provided that means are developed for launching and measuring the various types of spherical waves. In the absence of the measured values, the coefficients could be calculated numerically by solving the wave problem with the given antenna boundary condition; only the asymptotic approximation to the solution for the radiation field is required; the calculation of coupling into the feed-line is more complicated.

The method described here makes it possible to think about antenna problems of the type considered without explicit representation of the waves in space. It is equivalent to the circuit method applied to waves on a system of coupled transmission lines.

§ 6. *High Gain Antennas*

If we had to deal with a highly directive antenna such as an array, it would not be convenient to use spherical waves to represent the radiation because of the large number of them which must be superposed. We therefore employ plane waves as the basis of representation, and normalize the wave-functions to give unit power per unit area. The wave-functions are distinguished no longer by pairs of integral indices but by means of the direction of their propagation vector $\underline{\kappa}$ which is continuously variable in direction, although constant in length. The incident wave is represented as

$$I(r, \theta, \phi; \varpi) = \frac{1}{k^2}\iiint p(\underline{\kappa}; \varpi)\delta(k - \kappa)\, e^{j\underline{\kappa}\,\underline{r}} d\underline{\kappa}, \tag{58}$$

where $\delta(\ \ )$ is Dirac's δ-Function and integration is over the whole space spanned by the vector $\underline{\kappa}$. Within a constant factor, $p(\underline{\kappa}; \varpi)$ is the Fourier transform of $I(r, \theta, \phi; \varpi)$; the coordinate ϖ serves to distinguish polarization of the waves and will be defined more precisely in the sequel. As before, $\theta = 0$ is the direction of the incident beam, and if it is linearly polarized, $\theta = \dfrac{\pi}{2}$, $\phi = 0$ is the direction of the electric vector in it.

Let (ω, ψ) be the angular coordinates of the unit vector κ_1 referred to (θ, ϕ) as pole. Then

$$\left.\begin{aligned} &\cos\omega = \cos\theta\cos u + \sin\theta\sin u\cos(\phi - v) \\ &\sin\psi\sin\omega = \sin u\sin(\phi - v) \\ \text{and}\qquad &\sin u\, du\, dv = \sin\omega\, d\omega\, d\psi \end{aligned}\right\} \tag{59}$$

where (u, v) are the angular coordinates of $\underline{\kappa}_1$ referred to the pole of the (θ, ϕ) system.

Suppose that $p(\underline{\kappa})$ is a surface harmonic of nth order. We may write (Whittaker and Watson, 1915, § 18.61)

$$p_n(u, v) = A_n(\theta, \phi)P_n(\cos\omega) + \sum_{m=1}^{n} [A_n^{(m)}(\theta, \phi)\cos m\psi + B_n^{(m)}(\theta, \phi)\sin m\psi]\, P_n^{(m)}(\cos\omega), \tag{60}$$

and for the corresponding space-wave

$$\begin{aligned} I_n(r, \theta, \phi) &= A_n(\theta, \phi)\int_0^{2\pi}\int_0^{\pi} e^{jkr\cos\omega}P_n(\cos\omega)\sin\omega\, d\omega\, d\psi \\ &= 2\pi\, j^n\sqrt{\frac{\pi}{2kr}}\,[H^{(1)}_{n+\frac{1}{2}}(kr) + H^{(2)}_{n+\frac{1}{2}}(kr)]\,A_n(\theta, \phi). \end{aligned} \tag{61}$$

Now since

$$P_n(\cos\omega) = P_n(\cos\theta)P_n(\cos u) + 2\sum_{m=1}^{n}\frac{(n-m)!}{(n+m)!} P_n^{(m)}(\cos\theta)P_n^{(m)}(\cos u)\cos m(\phi - v) \tag{62}$$

and $$A_n(\theta, \phi) = \frac{2n+1}{4\pi}\int_0^{2\pi}\int_0^{\pi} p_n(u, v)P_n(\cos\omega)\sin\omega\, d\omega\, d\psi \quad (63)$$

we see that the expansion of $A_n(\theta, \phi)$ in terms of the surface harmonics $P_n(\cos\theta), P_n^{(m)}(\cos\theta)\,{\cos \atop \sin}\, m\phi$ has the same coefficients as that of $p(u, v)$ in terms of the harmonics where u, v replace θ and ϕ respectively. Consequently to each of the surface harmonics entering the representation $p(u, v)$ there corresponds an incoming and outgoing spherical wave.

From the form of I_n in equation (61), it is evident that it represents either rE_r or rH_r in an nth order spherical wave of T.M. or T.E. type. The transverse field components which interest us in considering radiation, are conveniently treated by means of Fourier transforms. The following relations follow from Maxwell's equations for a T.M. system. Let V denote rE and let **U** denote r**H**. Then (Frenkel, 1926; Stratton, 1941)

$$\left.\begin{aligned} \frac{\partial^2 V_\theta}{\partial r^2} + k^2 V_\theta &= \frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial V_r}{\partial \theta}\right) & \frac{\partial U_\theta}{\partial r} &= -\frac{jk}{Z_0}V_\phi \\ \frac{\partial^2 V_\phi}{\partial r^2} + k^2 V_\phi &= \frac{\partial}{\partial r}\left(\frac{1}{r\sin\theta}\frac{\partial V_r}{\partial \phi}\right) & \frac{\partial U_\phi}{\partial r} &= \frac{jk}{Z_0}V_\theta \end{aligned}\right\}. \quad (64)$$

The Fourier transforms of the differential operators in these equations are

$$\left.\begin{aligned} F\frac{\partial}{\partial r} &= j\kappa\cos\omega \\ F\frac{1}{r}\frac{\partial}{\partial\theta} &= j\kappa\sin\omega\cos\psi \\ F\frac{1}{r\sin\theta}\frac{\partial}{\partial\phi} &= j\kappa\sin\omega\sin\psi \end{aligned}\right\}. \quad (65)$$

Hence

$$\left.\begin{aligned} FV_\theta &= \frac{-\kappa^2\sin\omega\cos\omega\cos\psi}{k^2-\kappa^2\cos^2\omega}FV_r, & FU_\theta &= \frac{-k}{Z_0\kappa\cos\omega}FV_\phi \\ FV_\phi &= \frac{-\kappa^2\sin\omega\cos\omega\sin\psi}{k^2-\kappa^2\cos^2\omega}FV_r, & FU_\phi &= \frac{k}{Z_0\kappa\cos\omega}FV_\theta \end{aligned}\right\}. \quad (66)$$

Corresponding equations can be developed for T.E. waves. The only changes required are to substitute U for V, to change the sign of j in the second group of (64) and to transfer Z_0 to the numerator in these equations; obvious changes follow in (66).

We have chosen $\phi = 0$ to coincide with the electric vector in the incident plane wave: consequently in that wave

$$\begin{aligned} V_r &= \sin\theta\cos\phi\, I \\ Z_0 U_r &= \sin\theta\sin\phi\, I. \end{aligned} \qquad (67)$$

It follows therefore that the Fourier transforms of the meridianal and transverse components of electric force are given by

$$FE_\theta = \Theta FI$$

and (68)

$$FE_\phi = \Phi FI$$

where the operators Θ and Φ are

$$\Theta = \frac{-\kappa\sin w\sin\theta}{k^2 - \kappa\cos^2\omega}\left\{k\sin\psi\sin\phi + \kappa\cos\omega\cos\psi\cos\phi\right\}$$

$$\Phi = \frac{-\kappa\sin w\sin\theta}{k^2 - \kappa^2\cos^2\omega}\left\{\kappa\cos\omega\sin\psi\cos\phi - k\sin\phi\cos\psi\right\}. \qquad (69)$$

We have now to consider the representation of the scattered wave. We define ϖ as the angle between the electric vector of the linearly polarized plane wave (propagation vector $\underline{\kappa}$) and the intersection of $\phi = 0$ with the plane normal to $\underline{\kappa}$.

Since, whereas the incident radiation is a standing spherical wave system, the scattered wave is an outgoing spherical wave, the δ-Function contained as a factor in its Fourier transform must represent this fact. The desired result may be achieved by means of a generalized Fourier integral (Titchmarsh, 1937, p. 43). Split

$$\delta(k) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{jkt}\,dt$$

into two parts $\delta_+(k) + \delta_-(k)$ defined by

$$\delta_+(k) = \frac{1}{2\pi}\int_0^{\infty} e^{jkt}\,dt \qquad \delta_-(k) = \frac{1}{2\pi}\int_{-\infty}^{0} e^{jkt}\,dt. \qquad (70)$$

The former is represented by

$$\delta_+(k) = \frac{-1}{2\pi jk} \qquad (71)$$

on the complex k-plane, provided that $Im(k) > 0$. Heisenberg (1943) has shown that the factor $\delta_+(k - \kappa)$ in the Fourier transform corresponds to outgoing waves.

Thus we represent the scattered wave by

$$S(r, \theta, \phi; \varpi) = \frac{1}{k^2}\iiint s(\underline{\kappa}; \varpi)\,\delta_+(k - \kappa)\,e^{j\underline{k}\cdot\underline{r}}\,d\underline{k}. \qquad (72)$$

The system of coupling coefficients now becomes a function of the two vectors $\underline{\kappa}_1$, κ_0 and of the two polarization coordinates ϖ, ϖ_0 viz.

$$K(\underline{\kappa}_0, \underline{\kappa}_1; \varpi_0, \varpi_1)$$

and

$$s(\underline{\kappa}_0; \varpi_0) = \iiint \sum_{\bar{\omega}_1} K(\kappa_0, \underline{\kappa}_1; \varpi_0, \varpi_1)\, p(\kappa_1; \varpi_1)\, d\kappa_1, \tag{73}$$

or we may regard s, p, and K as matrices with continuously variable parameters describing the rows and columns, thus

$$s = Kp. \tag{74}$$

The incoming incident wave is represented by

$$p(\underline{\kappa}_1)\ \delta_-(k - \kappa_1) \tag{75}$$

and the outgoing by

$$\{p(\underline{\kappa}_0) + K(\underline{\kappa}_0, \underline{\kappa}_1; \varpi_0)\, p(\kappa_1)\}\ \delta_\perp(k - \kappa_0)$$

or

$$\{\delta(\underline{\kappa}_0 - \underline{\kappa}_1) + K(\underline{\kappa}_0, \underline{\kappa}_1; \varpi_0)\}\ p(\underline{\kappa}_1)\, \delta_\perp(k - \kappa_0). \tag{76}$$

Thus in place of the series of complex reflexion coefficients w_n we now have the matrix

$$w = \delta(\underline{\kappa}_0 - \underline{\kappa}_1) + K(\underline{\kappa}_0, \kappa_1; \varpi_0), \tag{77}$$

and we define the generalized wave-impedance as the matrix

$$-(\delta(\underline{\kappa}_0 - \underline{\kappa}_1) + 2K^{-1}) \tag{78}$$

analogous to

$$z = \frac{1 + w}{1 - w}.$$

In considering the large antenna structure necessary to give a well-directed beam, we expect that the spectrum of its resonant frequencies is much more closely packed than that of a small antenna. Accordingly it is unreasonable to assume that the function K can be obtained from the radiation pattern of the antenna corresponding to the mean frequency of the band over which the antenna is expected to operate. The radiation pattern will, in fact, be fairly strongly dependent on frequency. We may regard K as built up as follows:

$$K(\underline{\kappa}_0, \underline{\kappa}_1; \varpi_0, \varpi_1) = \sum_n \frac{a_n K_n(\underline{\kappa}_0, \kappa_1; \varpi_0, \varpi_1)}{\sqrt{(\kappa_n^2 - \kappa_1^2) + \kappa_1^2 b_n^2}} \tag{79}$$

where b_n is inversely proportional to the Q of the antenna at the nth resonant frequency $\frac{c\kappa_n}{2\pi}$. K_n is the function of the form given by the radiation pattern of the antenna for the nth resonant frequency.

In order to determine the effective cross-section of a high gain antenna in terms of the foregoing symbolism, it remains for us to find an expression for the scattered energy.

Consider a plane wave Σ polarized in the direction (ω, ψ) referred to (θ, ϕ) as pole, (ω', ψ') being the direction of the electric force in this linearly polarized wave. Then in terms of Fourier transforms

$$\begin{aligned} V_r &= \cos\omega' \Sigma \\ U_r &= \sin\omega' \sin\omega \sin(\psi' - \psi)\Sigma. \end{aligned} \tag{80}$$

Hence, making use of

$$\cos\omega \cos\omega' = -\sin\omega \sin\omega' \cos(\psi' - \psi),$$

we find for the operators Θ and Φ

$$\Theta = \frac{\kappa^2 \sin^2\omega \sin\omega'}{k^2 - \kappa^2\cos^2\omega}[\kappa \cos\psi \cos(\psi' - \psi) - k \sin\psi \sin(\psi' - \psi)]$$

$$\Phi = \frac{\kappa^2 \sin^2\omega \sin\omega'}{k^2 - \kappa^2\cos^2\omega}[\kappa \sin\psi \cos(\psi' - \psi) + k \cos\psi \sin(\psi' - \psi)]$$

or when $\kappa \to k$

$$\left.\begin{aligned} \Theta &\to \sin\omega' \cos\psi' \\ \Phi &\to \sin\omega' \sin\psi' \end{aligned}\right\}. \tag{81}$$

In the scattered wave, therefore,

$$E_\theta = \frac{1}{k^2}\iiint \Theta\, s(\underline{\kappa}; \varpi)\, \delta_+(k - \kappa)\, e^{j\kappa r\cos\omega} \sin\omega\, d\omega\, d\psi\, \kappa^2 d\kappa. \tag{82}$$

Since κ has a negative imaginary part, and we are concerned with r very large, we may apply in essence the method of steepest descent to integrate with respect to ω. The important part of the integral occurs for $\omega = 0$ and we obtain

$$E_\theta = \frac{2\pi}{k^2}\int_0^\infty \Theta\, s(\kappa; \varpi)\frac{e^{j\kappa r}}{j\kappa r}\kappa^2 \delta_+(k - \kappa)\, d\kappa \tag{83}$$

where the conversion of $\underline{\kappa}$ into κ indicates that $\underline{\kappa}$ is in the direction (θ, ϕ). Integrating with respect to κ on the complex plane, we have

$$E_\theta = \frac{e^{jkr}}{jrk}\sin\omega' \cos\psi'\, s(k; \varpi) \tag{84}$$

and

$$|E_\theta|^2 + |E_\phi|^2 = \frac{1}{k^2r^2}\sin^2\omega'\, |s(k; \varpi)|^2. \tag{85}$$

Since for transverse waves $\sin\omega' = 1$, the energy flux is

$$\frac{1}{2Z_0k^2}\iint |s(k; \varpi)|^2 \sin\theta\, d\theta\, d\phi. \tag{86}$$

Hence if $s(k; \varpi)$ is based on $p(\underline{\kappa}; \varpi) = \delta(\cos u - 1)\,\delta(v)$ where

$$d\underline{\kappa} = \kappa^2 d\kappa \sin u\, du\, dv,$$

corresponding to a linearly polarized plane wave of unit amplitude, the effective cross-section is

$$\frac{1}{k^2}\iint \sum_{\bar{\omega}} |s(k;\varpi)|^2 \sin\theta\, d\theta\, d\phi. \tag{87}$$

To the foregoing representation may be adjoined that of the system of waves localized at the antenna itself either within or outside the feed.

Saskatoon,
April 17, 1945.

REFERENCES

FRENKEL, J. Elektrodynamik (math. append.), Berlin. 1926.
HEISENBERG, W. Die "beobachtbaren Grossen" in der Theorie der Elementarteilchen. Zeits. f. Phys., 120: 513-38. 1943.
SCHELKUNOFF, S. A. Transmission Theory of Plane Electromagnetic Waves. Proc. Inst. Rad. Eng., 25: 1457. 1937.
——————— Transmission Theory of Spherical Waves. Trans. A.I.E.E., 57: 714. 1938.
SLATER, J. C. Microwave Transmission, McGraw-Hill. 1942.
SIRATTON, J. A. Electromagnetic Theory, McGraw-Hill. 1941.
TITCHMARSH, E. C. Theory of Fourier Integrals, Oxford. 1937.
WHITTAKER, E. T. and G. N. WATSON. Modern Analysis, Cambridge. 1915.

TRANSACTIONS OF

THE ROYAL SOCIETY OF CANADA

SECTION III

CHEMICAL, MATHEMATICAL, AND PHYSICAL SCIENCES

THIRD SERIES—VOLUME XL—SECTION III

MAY, 1946

56669

⁂

OTTAWA
THE ROYAL SOCIETY OF CANADA
1946

TABLE OF CONTENTS

Some Investigations in the Projective Differential Theory of Scrolls By C. T. SULLIVAN, F.R.S.C 1

The System: Iodine—Bromine—Chlorine, at 29.8°. By A. N. CAMPBELL, F.R.S.C, and L. W SHEMILT 17

A Generalised Integral. By R. D. JAMES, F.R.S.C., and WALTER H. GAGE 25

A Group Theoretic Characterisation of the General Projective Collineation Group. By N. S. MENDELSOHN 37

On a Boundary Value Problem for a Clamped Plate. By A. WEINSTEIN and J. A. JENKINS . 59

Transactions of the Royal Society of Canada

SECTION III

SERIES III MAY, 1946 VOL. XL

PRESIDENTIAL ADDRESS

SOME INVESTIGATIONS IN THE PROJECTIVE DIFFERENTIAL THEORY OF SCROLLS

By C T. SULLIVAN, F.R.S.C.

1. INTRODUCTION

IN the manuscript of a paper entitled "Scrolls Whose Asymptotic Curves are Cubics," I have shown that there are precisely six types of scrolls with cubic asymptotic curves If s denotes the degree of a scroll, and the symbols $S(1, 1, s)$, $S(\overline{1, 1}, s)$ indicate that it belongs to a non-special or a special linear congruence, then the possible types are:

$$S(1, 1; s),\ s = 4, 5, 6; \text{ and } S(\overline{1, 1}; s),\ s = 3, 4, 6.$$

Type $S(\overline{1, 1}; 3)$ is the Cayley cubic scroll, which has long been known to possess cubic asymptotic curves.

This manuscript has been submitted to several referees, and the willing scrutiny of well-versed private readers. Excepting the Cayley cubic scroll, no question has been raised as to the novelty of the results obtained or the originality of the methods employed. However, the impression prevailed and was expressed that the results might be obtainable by less formidable methods. These methods emerged as essential after the writer found that several claims in different theories concerning ruled surfaces and curves belonging to linear complexes are not quite exhaustive, being subject to exception. While no avenue of simplification has been suggested, some of the observations recorded appear to indicate that an unusual combination of methods peculiar to diverse theories rendered the exposition unnecessarily difficult.

It therefore becomes necessary to recast the work in two distinct parts. The first part deals with Projective Differential Theory[1] and its immediate geometrical contacts. The second part will be concerned with the process of generating the scrolls required, the algebraic-geometrical methods of segregating and exhibiting the types, and finally the allocation of these types within existing algebraic classifications, when available. Here we shall confine attention to the first of these

56669

parts, which provides among other results a constructive and critical basis for methods evolved in the second part.

Apart from a theorem of fundamental importance in both parts of the work, this paper deals with new material not touched upon in the original manuscript. The introduction here of this new material necessitates a slight increase in known details pertaining to the analytic basis of the theory employed. However, only the minimum essentials are included. The beautiful geometrical properties of an adjoined system of equations are omitted entirely, as not being imperative for the exposition in hand.

2. ANALYTIC-GEOMETRIC PRELIMINARY

The theory required now is based on the consideration of the system of differential equations:

$$\text{(A)} \qquad \begin{aligned} y'' + p_{11}y' + p_{12}z' + q_{11}y + q_{12}z &= 0, \\ z'' + p_{21}y' + p_{22}z' + q_{21}y + q_{22}z &= 0, \end{aligned}$$

where the coefficients are analytic functions of x in the vicinity of the value of x considered. According to a fundamental existence theorem in differential equation theory, the system (A) defines two functions y and z analytic near enough to the point $x = x_0$; and these functions can be made to satisfy the further condition that y, z, y', z' shall assume prescribed values at $x = x_0$. Such a system of solutions is called the *general boundary system of solutions* of the system (A).

Now let (y_i, z_i), $i = 1, \ldots, 4$ be any four systems of solutions of (A) such that the determinant $D = |y_1z_2y_3'z_4'|$ is not identically zero. Then by solving a set of linear algebraic equations it is possible to find c's such that

$$y = \sum_1^4 c_iy_i, \quad z = \sum_1^4 c_iz_i$$

form a system of general boundary solutions. The pairs of solutions (y_i, z_i) just considered are called a *fundamental system of simultaneous solutions*.

Conversely, if four pairs of functions (y_i, z_i) are such that $D = |y_1z_2y_3'z_4'| \neq 0$, then by solving a set of linear algebraic equations a system of equations (A) can be determined for which these functions constitute a simultaneous fundamental system of solutions.

Let us now regard $(y_1, \ldots, y_4)$ and $(z_1, \ldots, z_4)$ as the homogeneous coordinates of points P_y and P_z in space. If (A) is integrated these quantities are expressed as functions of x, and as x varies P_y and P_z describe two curves C_y and C_z, called guide curves. The line L_{yz}

joining P_y and P_z generates the integrating ruled surface or scroll of the system (A) as x varies. It is clear that transformations of the form, $\eta = \alpha y + \beta z$, $\zeta = \gamma y + \delta z$ enable us to convert the curves C_y and C_z into two other guide curves C_η and C_ζ on the scroll; but the generator corresponding to a given value of x remains undisturbed. A transformation of the form $\xi = f(x)$ changes the parametric representation of the curves without changing the curves or their point-to-point correspondence.

On the other hand, if D vanishes identically, it is possible to find four functions λ, μ, ν, ρ, such that

$$\lambda y_k + \mu y_k' + \nu z_k + \rho z_k' = 0, \quad k = 1, 2, 3, 4.$$

This may be written

$$(d/dx)(\mu y_k + \rho z_k) = (\mu' - \lambda)y_k + (\rho' - \nu)z_k.$$

But $\mu y_k + \rho z_k$ are the coordinates u_k of a point P_u on L_{yz}. The right-hand member gives a second point P_v on L_{yz} whose coordinates are v_k. The point P_u describes a curve C_u whose tangents, as the above equation shows, are the lines L_{yz}. The ruled surface is therefore a developable whose generators are the tangents to C_u. In particular, P_v may coincide with P_u. This will be so if $\mu' - \lambda = \omega\mu$, $\rho' - \nu = \omega\rho$, i.e., if $\mu'\rho - \mu\rho' = \lambda\rho - \mu\nu$. In this case P_u remains fixed, and the developable becomes a cone.

In the consideration of a fundamental system of solutions of (A), it may happen that the determinant D vanishes for some value, say $x = a$. The current geometrical interpretation of this phenomenon is: that in the vicinity of such a generator, the ruled surface resembles a developable; that two consecutive generators intersect at the *pinch point*; that the tangent plane is stationary along such a generator. In Cayley's terminology, the generator is said to be *torsal*.

However, it will be shown below that a singular generator of a type entirely different from this may occur at the point $x = a$, where D vanishes. It may therefore be affirmed that the actual organic character of this analytic phenomenon is not as yet capable of precise geometrical prediction.

3. TRANSFORMATIONS OF THE SYSTEM (A)

The determination of the asymptotic curves on an integral scroll corresponds to transforming (A) into *asymptotic form*;

$$(1) \qquad \begin{aligned} y'' + p_{11}y' + q_{11}y + q_{12}z &= 0 \\ z'' + p_{22}z' + q_{21}y + q_{22}z &= 0. \end{aligned}$$[1]

---

[1] E. J. Wilczynski, Projective Differential Geometry of Curves and Ruled Surfaces. Chapter V. We shall refer to this work as W.

The form of this system shows that C_y and C_z are asymptotic curves on the integrating ruled surface. Briefly, if in the equation of the osculating plane to C_y at P_y we substitute for second derivatives from the above system and reduce the resulting determinant, we obtain the equation of the tangent plane to the scroll at P_y. Similarly C_z is asymptotic on the scroll.

If we transform (1) by the substitutions

$$y = \eta e^{-\frac{1}{2}\int p_{11}dx}, \quad z = \zeta e^{-\frac{1}{2}\int p_{22}dx},$$

we obtain a new system of equations in η and ζ of the form:

$$\begin{aligned} y'' + q_{11}y + q_{12}z &= 0, \\ z'' + q_{21}y + q_{22}z &= 0. \end{aligned} \tag{2}$$

The parametric curves have not been disturbed by these substitutions. The system (2) is said to be in the *semi-canonical form*; its guide curves are asymptotic on the scroll.

The substitutions which lead almost directly from (A) to the semi-canonical form are:[2]

$$y = \alpha\eta + \beta\zeta, \quad z = \gamma\eta + \delta\zeta, \quad \alpha\delta - \beta\gamma \neq 0,$$

where (α, γ) and (β, δ) are two pairs of solutions of the system of equations,

$$\begin{aligned} \rho' + \tfrac{1}{2}(p_{11}\rho + p_{12}\sigma) &= 0 \\ \sigma' + \tfrac{1}{2}(p_{21}\rho + p_{22}\sigma) &= 0. \end{aligned}$$

It is not necessary to include here the intermediate system of equations arising in the transition from (A) to (2), since this system of equations will not be required in the subsequent work.

The most general transformations which leave the semi-canonical form invariant, i.e., which replace a pair of asymptotic guide curves by another pair of asymptotic guide curves, are:

$$\xi = \xi(x), \quad \eta = (ay + bz)\sqrt{\xi'}, \quad \zeta = (cy + dz)\sqrt{\xi'},$$

where $\xi(x)$ is an arbitrary function, and where a, b, c, d are constants such that $(ad - bc) \neq 0$.

In particular, if $a = d = 1$, $b = c = 0$, the system (2) becomes

$$\begin{aligned} \frac{d^2\eta}{d\xi^2} + \rho_{11}\eta + \rho_{12}\zeta &= 0 \\ \frac{d^2\zeta}{d\xi^2} + \rho_{21}\eta + \rho_{22}\zeta &= 0, \end{aligned}$$

where

$$\begin{aligned} \rho_{11} &= (\alpha'' + q_{11}\alpha)/(\xi'^2\alpha) \\ \rho_{22} &= (\alpha'' + q_{22}\alpha)/(\xi'^2\alpha), \quad \alpha = 1/\sqrt{\xi'}. \end{aligned}$$

---

[2] W., chapter IV, § 6.

Since $\xi(x)$ is at our disposal, we can make $\rho_{11} - \rho_{22} = 1$ by taking $\xi = \int (q_{11} - q_{22})^{\frac{1}{2}} dx$. This assumes that $q_{11} - q_{22} \neq 0$. We now show that there is no loss of generality in this assumption. In short, a scroll whose asymptotic curves belong to linear complexes has at most two linear directrices. If C_y and C_z be selected distinct from these, then $q_{12} q_{21} \neq 0$. If then (2) is transformed by the substitution $y = \eta + b\zeta$, $z = \zeta$, where b is an arbitrary constant, we obtain a new semi-canonical system for which $\bar{q}_{11} - \bar{q}_{22} = (q_{11} - q_{22}) - 2bq_{21} \neq 0$. The *canonical system* of equations which we shall use here is the system (2) with the additional condition $q_{11} - q_{22} = 1$.

Later in the work the equations of scrolls sometimes occur in the asymptotic form and not in the semi-canonical form. It is therefore necessary to know when a linear transformation

$$y = a\eta + b\zeta, \quad z = c\eta + d\zeta, \quad \Delta = ad - bc \neq 0,$$

with constant coefficients, leads from an asymptotic system to an asymptotic system. Applied to an asymptotic system, this transformation converts it into a new system in which

$$\bar{p}_{11} = (adp_{11} - bcp_{22})/\Delta, \quad \bar{p}_{12} = bd(p_{11} - p_{22})/\Delta,$$
$$\bar{p}_{21} = ac(p_{11} - p_{22})/\Delta, \quad \bar{p}_{22} = (bcp_{11} - adp_{22})/\Delta.$$

Hence in order that the new system may be in the asymptotic form, it is both necessary and sufficient that $p_{11} = p_{22}$. This result justifies certain operations employed later in the work.

4. SPACE CURVES

The projective differential geometry of a space curve can be based on the consideration of the differential equation

$$y^{(4)} + 4p_1 y^{(3)} + 6p_2 y'' + 4p_3 y' + p_4 y = 0.\text{[3]} \tag{3}$$

This is done in a manner analogous to that already employed in basing the study of ruled surfaces on the consideration of the system (A). For our present purpose, the most fundamental invariant of (3) is: $\theta_3 = P_3 - 3P_2'/2$, where $P_2 = p_2 - p_1' - p_1^2$, $P_3 = p_3 - p_1'' - 3p_1 p_2 + 2p_1^3$. The vanishing of θ_3 is the *necessary and sufficient condition that the corresponding curve should belong to a linear complex.*

THEOREM 1. *If the asymptotic curves of a scroll belong to linear complexes, then they are projectively equivalent.*

This theorem was proved in part in a former paper.[4] In order to avoid separate treatment of cases in which a certain invariant θ_4 of weight four does and does not vanish, we make use of the canonical system of equations defined above, which differs from that employed in the reference cited.

---

[3] W., pp. 238-55.

[4] C. T. Sullivan, Properties of Surfaces, etc. Trans. Amer. Math Soc., 15: 167-96. 1914.

The differential equations of C_y and C_z are formed by methods set forth on pages 168-70 of the paper already quoted. If we proceed from the canonical system defined here, and carry out the calculations indicated, then there will be found as follows:

C_y: $\theta_3 - P_3 - \frac{3}{2}P'_2 = \frac{1}{4}\{(q'_{11} - q'_{22}) - (q'_{12}/q_{12})(q_{11} - q_{22})\} = -\frac{1}{4}q'_{12}/q_{12}$.
Since the curve belongs to a linear complex, $q_{12} = c_{12}$. In a similar manner we find that $q_{21} = c_{21}$. The c's are constants such that $c_{12}c_{21} \neq 0$. The canonical system of equations for scrolls whose asymptotic curves belong to linear complexes is therefore:

$$
\begin{aligned}
y'' + q_{11}y + c_{12}z &= 0 \\
z'' + c_{21}y + q_{22}z &= 0
\end{aligned}
\tag{4}
$$

where $q_{11} - q_{22} = 1$ and the c's are as stated above.

For this system the invariant of weight four mentioned above has the value:

$$\theta_4 = I^2 - 4J = 16(1 + 4c_{12}c_{21}) = 16(1 + 4\nu).$$

If the scroll belongs to a special linear congruence $(1 + 4\nu)$ vanishes; but otherwise ν may have any value other than zero.

The differential equations of C_y and C_z as computed from (4) are:

$$
\begin{aligned}
C_y&: y^{(4)} + (q_{11} + q_{22})\, y'' + 2q'_{11}y' + (q''_{11} + q_{11}q_{22} - \nu)y = 0 \\
C_z&: z^{(4)} + (q_{11} + q_{22})\, z'' + 2q'_{22}z' + (q''_{22} + q_{11}q_{22} - \nu)z = 0.
\end{aligned}
$$

Now $q'_{11} - q'_{22} = q''_{11} - q''_{22} = 0$; and the validity of the theorem is established.

5. A Scroll Possessing a Double Generator

Among the scrolls mentioned in the introduction there occur some that possess one or two singular generators of a character that escapes detection by the methods of Projective Differential Theory hitherto available. Adequate geometrical interpretation of singularities among the coefficients of (A) is not yet available.

Singular generators of the type we have in mind may be illustrated in a fairly general way as follows:

Let us consider the guide curves C_y and C_z defined by the polynomials:

$$
\begin{aligned}
y^{(1)} &= 1 + a_1u + a_2u^2 + \dots = g_1(u), \\
y^{(2)} &= \phantom{1 + {}} b_1u + b_2u^2 + \dots = g_2(u), \\
y^{(3)} &= \phantom{1 + {}} c_1u + c_2u^2 + \dots = g_3(u), \\
y^{(4)} &= \phantom{1 + {}} d_1u + d_2u^2 + \dots = g_4(u), \quad \text{and}
\end{aligned}
\tag{5}
$$

$$
\begin{aligned}
z^{(1)} &= \alpha_1 u + \alpha_2 u^2 + \ldots = h_1(u),\\
z^{(2)} &= \beta_1 u + \beta_2 u^2 + \ldots = h_2(u),\\
z^{(3)} &= \gamma_1 u + \gamma_2 u^2 + \ldots = h_3(u),\\
z^{(4)} &= 1 + \delta_1 u + \delta_2 u^2 + \ldots = h_4(u),
\end{aligned}
$$

where $g_s(u)$ is of degree m_s and $h_s(u)$ is of degree n_s. The edge AD of the tetrahedron of reference is then a generator of the scroll. The coordinates $(g_i h_j)$ of an arbitrary generator L_{yz} are then found to be:

$$
\begin{aligned}
(6)\quad (12) &= \beta_1 u + (\beta_2 + a_1\beta_1 - b_1\alpha_1)u^2 + (\beta_3 + a_1\beta_2 + a_2\beta_1 - b_1\alpha_2 -\\
&\quad b_2\alpha_1)u^3 + \ldots,\\
(13) &= \gamma_1 u + (\gamma_2 + a_1\gamma_1 - b_1\alpha_1)u^2 + (\gamma_3 + a_1\gamma_2 + a_2\gamma_1 - c_1\alpha_2 -\\
&\quad c_2\alpha_1)u^3 + \ldots,\\
(14) &= 1 + (c_1 + \delta_1)u + (a_2 + a_1\delta_1 + \delta_2 - d_1\alpha_1)u^2 +\\
&\quad (a_3 + a_2\delta_1 + a_1\delta_2 + \delta_3 - d_1\alpha_2 - d_2\alpha_1)u^3 + \ldots,\\
(23) &= (b_1\gamma_1 - c_1\beta_1)u^2 + (b_1\gamma_2 + b_2\gamma_1 - c_1\beta_2 - c_2\beta_1)u^3 + \ldots,\\
(24) &= b_1 u + (b_2 + b_1\delta_1 - d_1\beta_1)u^2 + (b_3 + b_2\delta_1 + b_1\delta_2 -\\
&\quad d_1\beta_2 - d_2\beta_1)u^3 + \ldots\\
(34) &= c_1 u + (c_2 + c_1\delta_1 - d_1\gamma_1)u^2 + (c_3 + c_2\delta_1 + c_1\delta_2 -\\
&\quad d_1\gamma_2 - d_2\gamma_1)u^3 + \ldots.
\end{aligned}
$$

Consider an arbitrary line L through the point P $(\theta, 0, 0, 1)$ on AD; its coördinates are:

$$
\begin{aligned}
(7)\quad &(12) = \theta x_2,\ (13) = \theta x_3,\ (14) = \theta x_4 - x_1,\ (23) = 0,\ (24) = -x_2,\\
&(34) = -x_3.
\end{aligned}
$$

The values of u giving the intersections of L with the generators L_{yz} of the scroll are found from the equation,

$$
\begin{aligned}
(8)\quad &(12)(34)' + (12)'(34) + (13)(42)' + (13)'(42) + (14)(23)' + (14)'(23)\\
&= 0.
\end{aligned}
$$

By (6) and (7), (8) becomes,

$$
(9)\quad \{(\theta c_1 + \gamma_1)x_2 - (\theta b_1 + \beta_1)x_3\}u + \{\ldots\}u^2 + \{\ldots\}u^3 + \ldots = 0.
$$

If $b_1 = c_1 = \beta_1 = \gamma_1 = 0$, $u = 0$ is a double root of this equation, and AD is a double generator on the scroll. The coefficient of u^2 now becomes,

$$
\{(\theta c_2 + \gamma_2)\, x_2 - (\theta b_2 + \beta_2)x_3\}.
$$

If the line L lies in the plane,

$$
(10)\qquad (\theta c_2 + \gamma_2)x_2 - (\theta b_2 + \beta_2)x_3 = 0,
$$

then it meets the scroll in three points coincident with P. Thus the plane tangent to the scroll at P is given by equation (10). It follows that AD is a cuspidal generator of the torsal type, torsal being used here in the sense employed by Basset, *Geometry of Surfaces.*

It is to be observed nevertheless that if $u = 0$, the determinant

$$\begin{vmatrix} y^{(1)} & y^{(2)} & y^{(3)} & y^{(4)} \\ z^{(1)} & z^{(2)} & z^{(3)} & z^{(4)} \\ y^{(1)\prime} & y^{(2)\prime} & y^{(3)\prime} & y^{(4)\prime} \\ z^{(1)\prime} & z^{(2)\prime} & z^{(3)\prime} & z^{(4)\prime} \end{vmatrix}$$

vanishes, and the points P_y, P_z coincide with the vertices A and D respectively of the tetrahedron of reference. The double generator AD has not, however, the organic character predicted by Projective Differential Theory.

These results may now be confirmed by using point instead of line elements. Put $b_1 = c_1 = \beta_1 = \gamma_1 = 0$ in (5), and consider the values of u where the line,

(11) $\quad x_1 + ax_2 + bx_3 - \theta x_4 = 0, \; x_1 + Ax_2 + Bx_3 - \theta x_4 = 0,$

intersects the scroll. The equation giving these values is readily found to be of the form,

$$u^2 P_0(u) = 0.$$

If however the line lies in the plane (10), that is, if the first equation (11) is (10), then the values of u where it meets the scroll are readily found to be given by an equation of the form,

$$u^3 Q_0(u) = 0.$$

Here $P_0(u)$ and $Q_0(u)$ denote polynomials having a term independent of u.

6. A CORRESPONDENCE

If we represent a surface parametrically by the equations $x = x(u, v)$ etc., then in general when distinct values are assigned to one or other of the parameters, the corresponding curves on the surface are distinct. This may or may not be the case for scrolls belonging to a non-special linear congruence. The phenomenon to which attention is directed is comprised most conveniently in the following theorem.

THEOREM 2. *If a scroll belongs to a non-special linear congruence, then the correspondence between its asymptotic curves and the linear complexes that contain them may be either* (1, 1) *or* (2, 1).

This theorem is at variance with one of long standing in the literature pertaining to the geometry of non-special linear congruences.[5] In general it is improbable that the nature of the correspondence can be predicted *a priori* by theory emanating from the system of differential equations. However, the instances exhibited below show that it is possible to construct systems from which the nature of the correspondence is readily determined.

---

[5] W., p. 220 (Cremona, Halphen, Snyder).

The theorem will be proved when we show explicitly that both types are possible. We begin by taking the equations of the scroll in the canonical form (4). Taking cognizance of the known geometrical significance convolved in a certain quadratic covariant,[6] we refer the scroll to its linear directrices as guide curves by means of the transformation,

$$\eta = y - (1 + k)z/(2c_{21}), \quad \zeta = y - (1 - k)z/(2c_{21}),$$

where $k^2 = 1 + 4\nu$. The equations of the scroll now become

$$\begin{aligned} \eta'' + \tfrac{1}{2}(q_{11} + q_{22} - k)\eta &= 0 \\ \zeta'' + \tfrac{1}{2}(q_{11} + q_{22} + k)\zeta &= 0. \end{aligned} \tag{12}$$

The point $\xi = \eta + \lambda\zeta$, where λ is an arbitrary constant, describes an asymptotic curve on the surface as u varies, if we regard u as the independent variable in (12). The generators are u = constant.

Take the linear guide curves as opposite edges of the tetrahedron of reference, namely AB, $x_3 = x_4 = 0$, and CD, $x_1 = x_2 = 0$. The special complexes of the linear congruence to which the surface belongs then have the equations

$$C_1\colon\ \omega_{34} = 0 \text{ and } C_2\colon\ \omega_{12} = 0. \tag{13}$$

The two-term system containing the asymptotic curves is therefore $C_\mu = C_1 + \mu C_2 = 0$. The polar plane of the point P_ξ in the complex C_μ is,

$$\mu(\xi_2 x_1 - \xi_1 x_2) + (\xi_4 x_3 - \xi_3 x_4) = 0.$$

This plane is tangent to the scroll at P if $P_{\xi'}$ is on it, that is if,

$$\begin{aligned} &\mu\{(\eta_2 + \lambda\zeta_2)(\eta_1' + \lambda\zeta_1') - (\eta_1 + \lambda\zeta_1)(\eta_2' + \lambda\zeta_2')\} \\ +\ &\{(\eta_4 + \lambda\zeta_4)(\eta_3' + \lambda\zeta_3') - (\eta_3 + \lambda\zeta_3)(\eta_4' + \lambda\zeta_4')\} = 0. \end{aligned}$$

On using line coordinates this equation readily reduces to $\lambda^2(\zeta_3\zeta_4') + \mu(\eta_1\eta_2') = 0$. In virtue of (12), $\frac{d}{du}\{(\eta_1\eta_2')\} = \eta_1\eta_2'' - \eta_1''\eta_2 = 0$. Hence $(\eta_1\eta_2')$ = constant, and similarly $(\zeta_3\zeta_4')$ = constant. Clearly there is no loss of generality in assuming these constants to be unity; and we find,

$$\lambda^2 + \mu = 0.$$

It follows that the curves $+\lambda$ and $-\lambda$ belong to the same complex $C_1 - \lambda^2 C_2 = 0$. The correspondence is therefore (1, 1) or (2, 1) according as $\lambda = c$ and $\lambda = -c$ give rise to the same or distinct curves on the surface.

---

[6] W., pp. 122, 149.

7. A NON-ALGEBRAIC SCROLL WITH (2, 1) CORRESPONDENCE

If in (12) we put $q_{11}+q_{22}=0$ and $k=2$, then, in the language of kinematics, the system reduces to one consisting of an harmonic equation and an antiharmonic equation. The solutions of this system are immediately available in terms of elementary functions. The parametric equations of the scroll in cartesian form are then,

$$(14) \qquad x=\lambda\frac{\cosh u}{\cos u},\quad y=\lambda\frac{\sinh u}{\cos u},\quad z=\tan u.$$

From these we deduce immediately that

$$(15) \qquad xdy-ydx=\lambda^2\sec^2 u.du=\lambda^2 dz.$$

The curves $\lambda=c$ and $\lambda=-c$ belong to the same complex. The curves however are distinct and are not related in any way comparable to the branch concept of an algebraic function or its geometrical analogue in curve tracing.

The generators are given by the equations,

$$y=x\tanh u,\quad z=\tan u.$$

Hence the scroll lies within the space bounded by the planes $y=\pm x$, and consists of infinitely many vanes situated in separate compartments. The equation $\alpha=2\tan^{-1}(\tanh\pi/2)$ measures the angle of the compartment of greatest dihedral angle.

Consider the points $P(x_1, y_1, z_1)$ and $Q(x_2, y_2, z_2)$. In order that PQ may be a diameter of (15) the relations $x_1/x_2=y_1/y_2=1$ must hold. To be the principal diameter or axis, the points P and Q must lie on the z-axis. Thus the two-term system to which the asymptotic curves belong is co-axal. This result may also be inferred from the following more elementary considerations.

The equations of the tangent t_0 at any point

$$x_0=\lambda\frac{\cosh u_0}{\cos u_0},\quad y_0=\lambda\frac{\sinh u_0}{\cos u_0},\quad z_0=\tan u_0$$

on the asymptotic curve λ are,

$$\frac{x-x_0}{l'}=\frac{y-y_0}{m'}=\frac{z-z_0}{n'},\text{ where}$$

$$l'=\lambda(\cos u_0\sinh u_0+\sin u_0\cosh u_0)\div(1+\lambda^2 U_0)^{\frac{1}{2}},$$

$$m'=\lambda(\cos u_0\cosh u_0+\sin u_0\sinh u_0)\div(1+\lambda^2 U_0)^{\frac{1}{2}},$$

$n'=1\div(1+\lambda^2 U_0)^{\frac{1}{2}}$, $U_0=\cosh 2u_0+\sin 2u_0\sinh 2u_0$. If ψ is the angle between t_0 and the z-axis, and δ is the shortest distance between t_0 and the z-axis, then $\tan\psi=\lambda\sqrt{U_0}$, $\delta=\lambda/\sqrt{U_0}$. Hence $\delta\tan\psi=\lambda^2$.

The coordinates of the point N where δ meets t_0 are found to be

$$\alpha_0 = (\lambda/U_0)(\cos u_0 \cosh u_0 + \sin u_0 \sinh u_0),$$
$$\beta_0 = (-\lambda/U_0)(\cos u_0 \sinh u_0 + \sin u_0 \cosh u_0),$$
$$\gamma_0 = \tan u_0 - \frac{\sin u_0 \cosh 2u_0 + \cos u_0 \sinh 2u_0}{U_0 \cos u_0}$$

This is the point where t_0 touches the circular cylinder of radius δ and axis Oz.

Consider next the asymptotic curves $+\lambda$ and $-\lambda$ and the points $\pm u_0$ on these curves, such that $u_0 < \pi/2$. This means that for simplicity of geometrical descriptions, we confine our attention to the chief vane of the scroll lying in the compartment bounded by the planes

$$y = \pm x \tanh \left(\tfrac{\pi}{2}\right).$$

On the cylinder $C(0, \delta)$plot the points $P(\alpha_0, \beta_0, \gamma_0)$ and $Q(\alpha_0, -\beta_0, -\gamma_0)$. From these points draw perpendiculars to the axis and produce to meet the surface of the cylinder at P' and Q'. The tangents to the cylinder at P and Q that make an angle ψ with the axis are tangents to $+\lambda$ at the points indicated. Likewise similar tangents at P' and Q' are tangents to $-\lambda$.

8 THE $[p, q]$ SYSTEM OF EQUATIONS

To exhibit both types of correspondence by means of algebraic scrolls, we employ systems of differential equations in the asymptotic form with $p_{11} = p_{22} \neq 0$. The construction of suitable systems is most readily effected by means of the following consideration. If we modify Bessel's Equations slightly to the homogeneous form,

$$\frac{d^2y}{dx^2} + \frac{1}{x}\frac{dy}{dx} - \frac{n^2}{x^2}y = 0,$$

then the primary solutions obey the relation

$$y_n y_{-n} = 1.$$

We therefore direct our attention to the $[p, q]$ *system*,

$$(16)\qquad \frac{d^2y}{du^2} + \frac{1}{u}\frac{dy}{du} - \frac{p^2}{u^2}y = 0, \quad \frac{d^2y}{du^2} + \frac{1}{u}\frac{dy}{du} - \frac{q^2}{u^2}y = 0.$$

The corresponding scroll is represented in asymptotic parameters by the equations

$$(17)\qquad x_1 = \lambda u^p, \quad x_2 = \lambda u^{-p}, \quad x_3 = u^q, \quad x_4 = u^{-q}.$$

The algebraic equation of the scroll is then,

$$x_1^q x_4^p - x_2^q x_3^p = 0.$$

For irreducibility the integers p and q are subject to appropriate restrictions.

Suitable $[p, q]$ systems will next be selected to illustrate the correspondences (1, 1) and (2, 1), respectively. It is known that for a cubic scroll with quartic asymptotic curves the (1, 1) relation holds. This can be confirmed immediately by considering a [1, 2] system. Here for the purpose of illustration we shall confine our attention to two systems only, namely [1, 4] and [3, 5]. The first of these emerges as providing a quintic of type VI(A) in Edge's classification of quintic ruled surfaces.

8A. A QUINTIC SCROLL WITH ASYMPTOTIC CURVES OF DEGREE 8 AND (1, 1) CORRESPONDENCE

If we proceed from a [1, 4] system, then the cartesian equations of the scroll are found immediately to be:

$$x = \lambda u^5, \ y = \lambda u^3, \ z = u^8; \ x^4 - y^4 z = 0.$$

The plane $z = c$ meets the surface in the line at infinity in the xy plane and the four lines

$$z = c, \ x^4 - cy^4 = 0.$$

An arbitrary plane through the z-axis meets the surface in the generator; $x = cy$, $z = c^4$. It follows that the z-axis is a multiple directrix of order four, and the line at infinity in the xy plane is a simple directrix. The generators emerging from the points on the multiple directrix coincide at the origin and at the point infinity. These points may therefore be regarded as compound pinch points; and the generators through them possess similar peculiarities which we now consider. Apart from the pinch points it is clear that these generators possess no singularities in the nature of point multiplicity; otherwise, the line joining such a point to an arbitrary point on the multiple directrix would intersect the surface in more than five points. This fact also appears independently from the following consideration.

Let $(0, \beta, 0)$ be an arbitrary point on $0y$, and let $x/l = (y - \beta)/m = z/n = r$ be an arbitrary line through this point. The intersections of this line with the surface are given by the equation

$$F(\alpha, \beta, \gamma) + r\left(l\frac{\partial}{\partial\alpha} + m\frac{\partial}{\partial\beta} + n\frac{\partial}{\partial\gamma}\right)F + \ldots +$$
$$\frac{r^5}{\underline{|5}}\left(l\frac{\partial}{\partial\alpha} + m\frac{\partial}{\partial\beta} + n\frac{\partial}{\partial\gamma}\right)^5 F = 0,$$

where $\alpha = \gamma = 0$ and $F(x, y, z) = 0$ is the equation of the quintic. This equation reduces to

$$m^4 n r^5 - l^4 r^4 + n\beta^4 r = 0.$$

For tangency $n = 0$; the equation then has one root $r = \infty$ and a quadruple root $r = 0$. The generator under consideration has therefore only a tangent plane singularity.

Returning to the parametric representation we see that the asymptotic curves $\Gamma_8(\lambda)$ are of the eighth order; and that the point $(\lambda, -u)$ is identical with the point $(-\lambda, u)$. They lie on the cylinders $y^5 = \lambda^2 x^3$. It also follows directly that the complex containing $\Gamma_8(\lambda)$ is

$$\omega_{12} - 4\lambda^2 \omega_{34} = 0.$$

The correspondence between the curves Γ_8 and the complexes to which they belong is therefore (1, 1).

The null-planes of the points (0, 0, 0, 1) and (0, 0, 1, 0) in any complex of the pencil are $z = 0$ and $w = 0$ (plane at infinity), respectively. All curves Γ_8 pass through the compound pinch points. At each such point they have a common osculating plane, viz., the corresponding null-plane.

At the origin the common osculating plane has eight-point intersection with each curve. The common tangent $x = 0$, $z = 0$ has five-point intersection with each curve, while an arbitrary plane through the origin meets each curve in three points coincident with the origin. The numerical equality $3 + 5 = 8$ illustrates a property possessed by all asymptotic curves derived from $[p, q]$ systems.

8B. A SCROLL (OF THE EIGHTH DEGREE) WITH QUINTIC ASYMPTOTIC CURVES AND (2, 1) CORRESPONDENCE

If we consider a [3, 5] system, then the cartesian equations of the surface take the form

$$x = \lambda\theta^4,\ y = \lambda\theta,\ z = \theta^5;\ x^5 - y^5 z^3 = 0,\ \theta \geq 0.$$

The plane $z = c$ intersects the surface in five generators emerging from the z-axis and the line at infinity in the xy plane. The plane $x = cy$ intersects the surface in three generators emerging from the line at infinity in the xy plane and the z-axis. It follows that the z-axis is a multiple directrix of the fifth order and the line at infinity in the xy plane is a multiple directrix of the third order. The generators emerging from each of the points zero and infinity on the z-axis coincide. Likewise those emerging from the triple directrix coincide for the points at infinity on the x-axis and y-axis respectively. The faces of the coordinate tetrahedron (cartesian) intersect the surface according to the following scheme:

	Face	*Threefold*	*Fivefold*
1.	$x = 0$	$z = 0$	$y = 0$
2.	$y = 0$	$w = 0$	$x = 0$
3.	$z = 0$	$w = 0$	$x = 0$
4.	$w = 0$	$z = 0$	$y = 0$

Here $w = 0$ is of course the plane at infinity.

The generator joining the points $(0, 0, 0)$ and $(0, \infty, 0)$ is a multiple generator of order three. This may be seen independently in the following manner. An arbitrary line through $(0, \beta, 0)$ meets the surface where,

$$m^5n^3r^8 + A_1n^3r^7 + A_2n^3r^6 + (A_3n^3 - l^5)r^5 + A_4n^3r^4 + \beta^5n^3r^3 = 0,$$

where the A's do not contain n. Hence the generator is multiple of order three. The tangent plane along this generator meets the surface in five consecutive generators. A similar statement applies to the generator joining $(0, 0, \infty)$ and $(\infty, 0, 0)$.

The asymptotic curves $\Gamma_5(\lambda)$ are of order five, and lie on the cylinders, $y^4 = \lambda^4 x$. The curves $\Gamma_5(+\lambda)$ and $\Gamma_5(-\lambda)$ are distinct; but they belong to the same complex $5\omega_{12} - 3\lambda^2\omega_{34} = 0$. The correspondence in this case is therefore (2, 1).

All the curves $\Gamma_5(\lambda)$ pass through the two singular points on the z-axis; and they have at each of these points a common osculating plane. These planes are the null-planes $z = 0$, $w = 0$ determined by the complexes. The osculating plane $z = 0$ meets each Γ-curve in five points coincident with the origin. The tangent at the origin meets each curve in four points at the point of tangency. An arbitrary plane through the origin meets each curve in one point there. This further illustrates a property of Γ-curves derived from a $[p, q]$ system.

A $[p, q]$ system possesses considerable practical value. By means of such a system it becomes possible to write down the equations of a rational curve Γ_n of degree n that belongs to a linear complex, and the equation of the complex to which it belongs. The numerical relation to which attention was directed in the previous examples is a special case of the following more general result.

If at a point P on a Γ-curve determined by a $[p, q]$ system, the curve has σ, τ, ρ consecutive points in common with the osculating plane, the tangent, and an arbitrary plane through P, respectively, then the relation $\sigma = \tau + \rho$ holds. Consider a $[p, q]$ system, $q > p$, and let $\sigma = 2q$, $q + p = \tau$, $q - p = \rho$. The cartesian equation of the surface in parametric form may be taken as,

$$x = \lambda u^\tau, \ y = \lambda u^\rho, \ z = u^\sigma.$$

If now we identify P with the origin, the intersections of the entities specified with any $\Gamma(\lambda)$ are found directly to be the numbers indicated. We also see that,

$$\sigma = 2q = (q + p) + (q - p) = \tau + \rho.$$

This is the relation we were required to prove. It appears probable that this property persists at a singularity on any curve belonging to a linear complex.

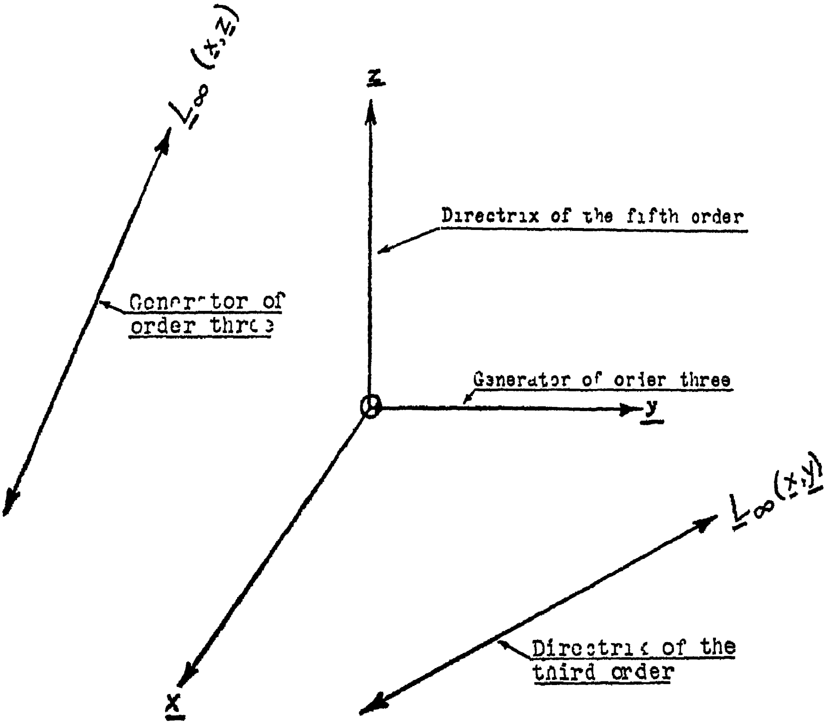


FIGURE 1. Singularities on the scroll defined by a [3, 5] system. The null-plane $z = 0$ intersects the surface in five consecutive generators along Oy. The null-plane $w = 0$ intersects the surface in five consecutive generators along $L_\infty(x, z)$.

THE SYSTEM: IODINE—BROMINE—CHLORINE, AT 29.8°

By A. N. CAMPBELL, F.R.S.C., and L. W. SHEMILT

1. INTRODUCTION

THE object of the present investigation was to establish whether or not ternary compounds of the halogens, such as $ICIBr_2$ and ICl_2Br, exist The literature of the binary halogen compounds is important and extensive and it is given in the bibliography. No such complete bibliography is to be found elsewhere and it may therefore be of some importance. The following binary compounds are known to exist, iodine monochloride, ICl, and iodine trichloride, ICl_3, which are well-known crystalline compounds. The monochloride exists in two forms, the metastable β melting at 13.9° and the stable α melting at 27.2°. The trichloride has a melting point of 101°. These compounds do not form solid solutions with one another, nor with iodine, nor, so far as is known, with chlorine.

There seems no reasonable doubt of the existence of a monochloride of bromine, melting, however, around −66° and therefore not a possible solid phase. The existence of a trichloride of bromine has not been claimed.

The existence of iodine monobromide, IBr, melting at 41°, is generally admitted. This compound forms a continuous series of mix-crystals with both iodine and bromine. In the temperature composition diagram, the solidus touches the liquidus at a point corresponding to the composition IBr. It is very doubtful whether a tribromide of iodine exists, although its existence as an ion in solution is still occasionally claimed.

As the temperature of this investigation (29.8°) was above the melting point of either of the forms of the monochloride of iodine, the binary solid phases to be expected were iodine (M.P. = 113.5°), iodine trichloride, and solid solutions of iodine in iodine monobromide.

Although all the component binary systems have been investigated by the method of thermal analysis, no work has been done on the ternary system. The nearest approach to it is a study by Plotnikow and Rokotjan (2) on the electrical conductivity of bromine solutions of iodine and iodine trichloride. In this work the authors made determinations of the solubility of iodine and of iodine trichloride in bromine.

The isothermal method was adopted by us, despite the obvious experimental difficulties, in order that we might use Schreinemakers' method of the wet "rest" and thus evaluate, without ambiguity, the composition of any ternary compounds which might exist under the conditions of the work

2. EXPERIMENTAL

Purity of materials. The iodine and bromine were of the highest purity commercially obtainable. The chlorine was obtained from cylinders and was purified by washing through distilled water and saturated copper sulphate solution (to remove any trace of hydrogen chloride), and finally dried by passage through concentrated sulphuric acid.

Apparatus. The thermostat was electrically controlled at a temperature of 29.8 ± 0.05°. Equilibrium was attained by stirring in a glass vessel fitted with a condenser. The liquid and solid phases were separated by filtration through a sintered glass filter, into a weighed container. The filtration was rapid so that alteration in composition due to loss of vapour was slight.

Procedure. On the basis of Terwogt's results (5), iodine and bromine were weighed out in such proportions as would give a solid phase and dried chlorine then bubbled in to a desired increase in weight. The total weight of the system varied between 100 and 250 grams. In order to attain equilibrium rapidly, the system was melted to a homogeneous liquid by raising the temperature to 40-50°. The container was then placed in the thermostat and the contents stirred for at least twelve hours.

Solutions in equilibrium with the trichloride (there is no solution simultaneously in equilibrium with trichloride and iodine or solid solutions of bromine and iodine) could not be prepared in the above way because it is difficult to get the solution to absorb sufficient chlorine We therefore prepared iodine trichloride by the method of Booth and Morris (1). Solutions were prepared from this by adding bromine and iodine or chlorine. It was not possible to prepare any solution containing more than 30 per cent chlorine, because the vapour pressure of chlorine then exceeds one atmosphere.

Method of analysis. Weighed samples for analysis were plunged in sodium hydroxide solution. Some care must be taken in this operation to ensure rapid immersion or there may be a loss of weighed sample by evaporation. An aliquot portion was then boiled with zinc dust for twenty minutes, then acidified with acetic acid and boiled for another twenty minutes. This procedure is sufficient to reduce the hypohalites: mineral acids tend to liberate free bromine or iodine.

The three halide radicles in solution were simultaneously determined potentiometrically. The method is described by Stock and Stahler (3) for the determination of chloride and iodide. Since the solubility products of silver iodide, silver bromide, and silver chloride are respectively of the orders 10^{-16}, 10^{-13}, and 10^{-10}, the method is capable of extension to the estimation of the three halogen ions. The presence of barium nitrate is necessary to prevent the adsorption, and consequent removal from solution, of bromide and chloride ion by the precipitated silver iodide. The attainable accuracy of the method was tested on solutions of known concentration and typical results are given in Table I.

TABLE I

Potentiometric Determinations of Halide in Solution

(Results are given in c.cs of silver nitrate solution)

	No.	*Calculated*	*Found*	*Per cent error*
I′	1	19.98	19.99	+ 0.05
	2	39.96	40.05	+ 0.2
	3	24.97	24.96	− 0.04
Br′	1	20.02	19.93	+ 0.5
	2	40.04	40.30	+ 0.6
	3	25.03	24.00	− 0.2
Cl′	1	19.98	20.18	+ 1.0
	2	2.00	2.20	+ 10.0
	3	2.00	2.28	+ 14.0

The results show that when the three ions are present in approximately equivalent amounts an accuracy well within 1 per cent is attainable, but, with small amounts of chloride, the error in the chloride determination is considerably increased. The estimation is undoubtedly most accurate for the iodide determination (the change of potential is sharpest and greatest here) and, since the quantity of iodine was never less than 30 per cent in any analysis, the estimations of iodine are almost free from error. An over-all picture of attainable accuracy resulted from adding the total weights of each halogen and comparing this with the weight of sample. This was done in every case. The results vary between 100 and 103 per cent of the original weight. It seems that the analytical procedure tends to give high results. The method, therefore, cannot be said to be of the highest accuracy and this lack of accuracy is shown by the slope of the tie-lines in the equilibrium diagram; the difficulty of estimating the three halogens in presence of one another, however, is notorious. Any purely chemical method would have given a worse result. Moreover, the method was sufficiently accurate for our purpose, since the results show without ambiguity the non-existence of ternary compounds.

Attainment of equilibrium All systems were left in the thermostat for at least twenty-four hours, i.e., twelve hours after cessation of stirring. That equilibrium was attained in this relatively short time was proved by approaching equilibrium from both sides, as follows A test system was divided into two parts. One part was heated until it was entirely liquid, then placed in the thermostat. The other part was cooled until it was almost solid, then stirred in the thermostat The analyses of the two tests agreed within the limits of experimental error

Equilibrium data. The analyses of saturated solutions and equilibrium wet "rests" are given in Table II.

TABLE II

	Composition of Liquid Phase (*Weight per cent*)			Composition of Wet "Rest" (*Weight per cent*)			
	I	Br	Cl	I	Br	Cl	*Nature of solid phase*
1	51.45	48.55	—	56.59	43.41	—	I ≈ Br Mix-crystal
2	50.24	46.85	2.91	55.34	42.66	2.00	"
3	50.79	46.01	3.20	55 40	42.12	2.48	"
4	51.52	44.56	3.92	56.20	41.71	2.09	"
5	59.24	34.73	6.03	61.20	35.37	3.43	"
6	63.80	29.27	6.93	64.70	31.07	4.23	"
7	66.11	26.50	7 30	66.35	28.65	5.00	"
8	71.80	19.77	8.43	72.43	21 06	6.51	"
9	74.42	15.44	10.14	76.00	16.65	7.35	"
10	77.76	11.29	10.95	81.00	12.04	6 95	"
11	81.23	7.67	11.10	87.45	6.99	5.61	"
12	83.07	4.49	12.44	89.59	3.81	6 60	"
13	84.72	—	15.28	93.23	—	6.79	Iodine
14	72.60	—	27.40	65.30	—	34.68	ICl_3
15	69.91	4.04	26.05	62.92	2 58	34.50	"
16	64.63	11.89	23.48	60.32	6.27	33.41	"
17	61.94	14.96	23.10	55.58	7.29	37.16	"
18	60.11	16.34	23.55	56.03	8.11	35.86	"
19	56.35	20.72	22.93	54.45	12 22	33.33	"
20	36.98	40.80	22.22	48.43	18.87	32.70	"
21	35.18	42.71	22.11	48.93	13.50	37.57	"
22	22.52	61.52	15.96	36.13	32.82	31.05	"

The above results are expressed graphically in Fig. 1.

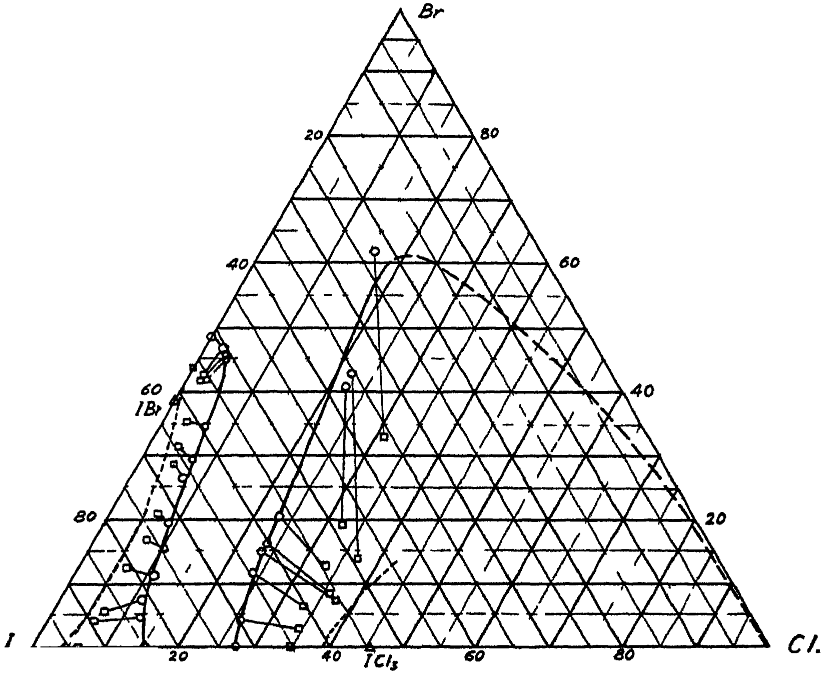


FIGURE 1 —The system Iodine—Bromine—Chlorine: 29.8° isothermal

3. DISCUSSION OF RESULTS

The systems investigated fall into two groups. The first thirteen, as can be seen from Fig. 1, represent solutions in equilibrium with solid solutions of iodine monobromide and iodine. There is no break in this curve, indicating, in agreement with Terwogt (5), that the series of solid solutions is continuous. Some comparison of the analytical results is possible with the thermal analytical results of Terwogt and the direct determinations of Plotnikow and Rokotjan (2): the comparison yields general agreement. The last point on the curve, referring to a system containing no bromine, is in good agreement with the results of Stortenbeker (4) for the system: iodine-chlorine. The analyses of solid phases when plotted are some distance from the I-Br side of the diagram. This, while it may be due to some extent to imperfect separation of solid phase from adhering mother liquor, does seem to indicate a certain solid solubility of chlorine in solid solutions of iodine and iodine bromide. Bromine being isomorphous with chlorine and with iodine bromide may indicate that chlorine is also isomorphous with iodine bromide. The maximum solid solubility, on this assumption, is about 4.5 per cent chlorine.

The second group of solutions is that of those in equilibrium with ICl_3 as solid phase. The diagram shows that at 29.8° there is no invariant solution saturated with both iodine and ICl_3. Neither is such a solution possible at lower temperatures since then ICl comes into being. On joining by a straight line the point on the base of the triangle representing ICl_3 to the apex of the triangle this series of solutions is separated into two, viz. those containing more iodine than ICl_3 and those containing more chlorine. Owing to the high vapour pressure of chlorine, it was impossible in an open apparatus to prepare any solutions of this latter series. The right-hand branch of the curve is therefore drawn in dotted in a purely hypothetical manner, except that the extreme right-hand point, representing a system containing no bromine, is obtained from the results of Stortenbeker (4). The results for the left-hand branch of the curve are in good agreement, as far as a comparison is possible, with those of Stortenbeker (4) and Plotnikow and Rokotjan (2). The tie-lines all run, approximately, in the direction of ICl_3, indicating that this alone is the solid phase, although here, too, the possibility of some solid solution is not excluded.

The results of this work show that, at 29.8°, the only solid phases existing are IBr, solid solutions of iodine and IBr (possibly containing chlorine), and ICl_3. There is no indication of the existence of ternary compounds of iodine, chlorine, and bromine. A careful search of the literature also failed to reveal any claim for the existence of such compounds as $IClBr_2$ or ICl_2Br.

One can speculate at length, in the light of modern theories of valence and electronic structure, electron affinity, etc., as to why these intermediate compounds do not exist, but to us the important matter seems to be that they do not.

4. SUMMARY

The ternary system: iodine–chlorine–bromine, has been investigated isothermally at 29.8°C. and the diagram for solid-liquid equilibrium obtained. No ternary compounds were found to exist in the solid state at 29.8°. It was verified that iodine forms a continuous series of solid solutions, possibly also containing chlorine, with iodine monobromide. The method of analysis used in this work was a potentiometric one, which has been extended to the estimation of the three halides.

BIBLIOGRAPHY OF BINARY HALOGEN COMPOUNDS

1. *Iodine–Chlorine*

Karsten. Z. anorg. Chem., **53-54**: 365. 1907.
Stortenbeker. Rec. Trav. chim., **7**: 152. 1888;
Z. physikal. Chem., **3**: 11. 1889.
Walden. Z. anorg. Chem., **68**: 307. 1910.

2. *Bromine–Chlorine*

Anwar-Ullah. Jour. Chem. Soc., 1176. 1932.
Balard. Ann. Chim. Phys., **32**: 371. 1826.
Barratt and Stein. Proc. Roy. Soc., **A 122**: 582. 1929.
Bornemann. Lieb. Ann., **189**: 209. 1877.
Butler and McIntosh. Proc. and Trans. Nova Scotia Inst. Sci., **17**: 23. 1927.
Karsten. Z. anorg. Chem., **53-54**: 365. 1907.
Lebeau. Comptes rendus, **143**: 589-91. 1906.
Loewig. Bromine and its Chemical Compounds. Heidelberg. 1829.
Lux. Ber., **63B**: 1156-8. 1930.
Meyer In Gmelin's Handbuch der anorganischen Chemie. 8th ed., System No. 7, p. 341.
Morgan and Burstall. Modern Developments in Inorganic Chemistry. P. 191. W. Heffer & Sons. 1936.
Taylor and Forscey. Jour. Chem. Soc., 2272. 1930.
Thomas and Dupuis Comptes rendus, **143**: 282. 1906.

3. *Bromine–Iodine*

Balard. Ann. Chim. Phys., **32**: 371. 1826.
Bodenstein and Schmidt. Z. physikal. Chem., **123**: 30. 1926.
Bornemann. Lieb. Ann., **189**: 204. 1877.
Lieb. Ann., **189**: 209. 1877.
Finkelstein. Z. physikal. Chem., **124**: 285. 1926.
Forbes and Faull. Jour. Amer. Chem. Soc., **55**: 1820. 1933.
Kruyt and Helderman. Z. physikal. Chem., **93**: 89. 1918.
Lagermark. Ber. d. Chem. Ges., **7**: 907. 1874.
Loewig. Pogg. Ann., **14**: 485. 1832.
McMorris and Yost. Jour. Amer. Chem. Soc., **53**: 2625. 1931; **55**: 552. 1933.
Plotnikov and Rokotjan. Z. physikal. Chem., **84**: 367. 1913.
Reade. Jour. Chem. Soc., **125**: 148. (1924).
Ruer. Z. physikal. Chem., **59**: 15. 1907.
Terwogt. Z. anorg. Chem., **47**: 203. 1905.
Wright. Jour. Chem. Soc., **109**: 1134. 1916.
Zeise. Z. Elektrochem., **41**: 267. 1935.

REFERENCES

(1) Booth and Morris. Inorganic Syntheses. Vol. 1, p. 166. McGraw-Hill. 1939.
(2) Plotnikov and Rokotjan. Jour. Russ. Phys. Chem. Soc., **47**: 723. 1915.
(3) Stock and Stähler. Quantitative Chemical Analysis. P.160. McGraw-Hill.1935.
(4) Stortenbeker. Z. physikal. Chem., **3**: 11. 1889.
(5) Terwogt. Z. anorg. Chem., **47**: 203. 1905.

A GENERALIZED INTEGRAL

By R. D. JAMES, F.R.S.C., and WALTER H. GAGE

1 INTRODUCTION

THIS paper defines what may be called a Perron second integral, or simply a P^2-integral, and establishes some of its properties. Although the definition and many of the proofs are similar to those of the ordinary Perron integral (Saks, McShane), the P^2-integral, as the example of §8 shows, is more general. It is probable that the P^2-integral will prove useful in the theory of trigonometric series

2. NOTATION AND PRELIMINARY DEFINITIONS

Let

$$\Delta_h^2 g(x) \equiv g(x+h) - 2g(x) + g(x-h).$$

Then

$$\overline{D}^2 g = \overline{\lim_{h\to 0}} \frac{\Delta_h^2 g(x)}{h^2}$$

and

$$\underline{D}^2 g = \underline{\lim_{h\to 0}} \frac{\Delta_h^2 g(x)}{h^2}$$

are the *upper and lower generalized second derivatives*, respectively, of $g(x)$. If $\overline{D}^2 g = \underline{D}^2 g$, the common value is denoted by $D^2 g$ and called the *generalized second derivative of* $g(x)$.

If $\Delta_h^2 g(x_0) = o(h)$ as $h \to 0$, the function $g(x)$ is said to be smooth at $x = x_0$.

If $g(x)$ has a right-hand derivative $g'_+(x)$, let

$$\Psi_h g(x) = 2\,\{g(x+h) - g(x) - hg'_+(x)\}$$

Then

$$Zg = \overline{\lim_{h\to 0}} \frac{\Psi_h g(x)}{h^2},$$

$$Z^+ g = \overline{\lim_{h\to 0+}} \frac{\Psi_h g(x)}{h^2},$$

and

$$Z^- g = \overline{\lim_{h\to 0-}} \frac{\Psi_h g(x)}{h^2},$$

are the *upper, upper right, and upper left modified second derivatives*, respectively, of $g(x)$.

A function $g(x)$ is *convex* in an interval (a, c) if

$$g(x_2) \leqq \frac{x_3 - x_2}{x_3 - x_1}\, g(x_1) + \frac{x_2 - x_1}{x_3 - x_1}\, g(x_3) \tag{2.1}$$

for all x_1, x_2, x_3 such that $a \leqq x_1 < x_2 < x_3 \leqq c$.

3. PROPERTIES OF CONVEX FUNCTIONS

Some of the results of this section are well known, but others appear to be new. Theorems 1 and 2 are fundamental.

THEOREM 1. *If $g(x)$ is continuous and smooth in (a, c) and if $\overline{D}^2 g \geqq 0$ for all x in (a, c) with the possible exception of a denumerable set, then $g(x)$ is convex.* (Hardy and Rogosinski, Theorem 95)

THEOREM 2. *If $g(x)$ is convex in an interval (a, c) and bounded above in some interval interior to (a, c), then $g(x)$ has the following properties:*

(a) *$g(x)$ is continuous in (a, c);*

(b) *the left and right-hand derivatives $g'_-(x)$ and $g'_+(x)$ exist for all x in (a, c);*

(c) *$g'_-(x) \leqq g'_+(x)$ for all x in (a, c);*

(d) *both $g'_-(x)$ and $g'_+(x)$ increase with x;*

(e) *the derivative $g'(x)$ exists for all x in (a, c) with the possible exception of a denumerable set;*

(f)
$$g'_+(x_1) \leqq \frac{g(x_3) - g(x_2)}{x_3 - x_2} \leqq g'_-(x_4),$$
$$\text{for } a \leqq x_1 \leqq x_2 < x_3 \leqq x_4 \leqq c.$$

(Hardy, Littlewood, and Polya, §3.18)

THEOREM 3. *Let $g(x)$ be continuous and convex in an interval (a, c), and let (α, γ) be any interval interior to (a, c). If $\overline{E}_k$ is the set of points in (α, γ) for which $\overline{Z}g \geqq k$, then*

$$\frac{g(c) - g(\gamma)}{c - \gamma} - \frac{g(\alpha) - g(a)}{\alpha - a} \geqq \frac{1}{4}\, k m_e(\overline{E}_k)$$

Proof: Let E_k^+ be the set of points in (α, γ) for which $Z^+g \geqq k$ and suppose that $k > 0$, $m_e(E_k^+) > 0$. If ϵ is any positive number and x any point in E_k^+, there exists an $h > 0$ such that

$$\frac{\Psi_h g(x)}{h^2} \geqq Z^+g - \epsilon \geqq k - \epsilon. \tag{3.1}$$

Thus every point in E_k^+ is the left-hand end-point of one or more intervals $(x, x + h)$ for which the inequality (3.1) holds. Hence, by a lemma of Sierpinski (Titchmarsh, §11.41), there is a finite non-overlapping set S of intervals $(x_\nu, x_\nu + h_\nu)$, $\nu = 1, \ldots, p$, such that

$$\frac{\Psi h_\nu g(x_\nu)}{h_\nu^2} \geqq k - \epsilon \text{ and } \sum_{\nu=1}^{p} h_\nu > m_e(E_k^+) - \epsilon$$

Then

$$\sum_{\nu=1}^{p} \left\{ \frac{g(x_\nu + h_\nu) - g(x_\nu)}{h_\nu} - g'_+(x_\nu) \right\}$$

$$\geqq \frac{1}{2}(k - \epsilon) \sum_{\nu=1}^{p} h_\nu > \frac{1}{2}(k - \epsilon)(m_e(E_k^+) - \epsilon) \tag{3.2}$$

Also, since $g(x)$ is convex, it follows from Theorem 2 and (2.1) that

$$\left\{ \begin{aligned} g'_-(x_1) - \frac{g(a) - g(\alpha)}{a - \alpha} &\geqq 0, \\ g'_+(x_1) - g'_-(x_1) &\geqq 0, \\ \left. \begin{aligned} g'_-(x_{\nu+1}) - \frac{g(x_\nu + h_\nu) - g(x_\nu)}{h_\nu} &\geqq 0 \\ g'_+(x_{\nu+1}) - g'_-(x_{\nu+1}) &\geqq 0 \end{aligned} \right\} & \ \nu = 1, \ldots, p - 1, \\ \frac{g(c) - g(\gamma)}{c - \gamma} - \frac{g(x_p + h_p) - g(x_p)}{h_p} &\geqq 0. \end{aligned} \right. \tag{3.3}$$

The addition of all the inequalities (3.2), (3.3) gives

$$\frac{g(c) - g(\gamma)}{c - \gamma} - \frac{g(\alpha) - g(a)}{\alpha - a} > \frac{1}{2}(k - \epsilon)(m_e(E_k^+) - \epsilon).$$

But ϵ is arbitrary and hence

$$\frac{g(c) - g(\gamma)}{c - \gamma} - \frac{g(\alpha) - g(a)}{\alpha - a} \geqq \frac{1}{2} k m_e(E_k^+). \tag{3.4}$$

Since $g(x)$ is convex, this inequality is trivial if $k \leqq 0$ or if $m_e(E_k^+) = 0$. Thus (3.4) remains true without restriction on k or $m_e(E_k^+)$.

Similarly, if E_k^- is the set of points in (α, γ) for which $Z^- g \geqq k$ then

$$\frac{g(c) - g(\gamma)}{c - \gamma} - \frac{g(\alpha) - g(a)}{\alpha - a} \geqq \frac{1}{2} k m_e(E_k^-) \tag{3.5}$$

Finally, $\bar{E}_k = E_k^+ + E_k^-$ and $m_e(\bar{E}_k) \leqq m_e(E_k^+) + m_e(E_k^-)$, so that the required result follows from (3.4) and (3.5).

Theorem 4. *Let $g(x)$ be continuous and convex in an interval (a, c) and let (α, γ) be any interval interior to (a, c). If E_k is the set of points in (α, γ) for which $\bar{D}^2 g \geqq k$, then*

$$\frac{g(c) - g(\gamma)}{c - \gamma} - \frac{g(\alpha) - g(a)}{\alpha - a} \geqq \frac{1}{4} k m_e(E_k).$$

Proof. The identity

$$\Delta_h^2 g(x) \equiv \tfrac{1}{2}\ \{\Psi_h g(x) + \Psi_{-h} g(x)\}$$

and the fact that $\Psi_h g(x) \geqq 0$, $\Psi_{-h} g(x) \geqq 0$ show that

$$\overline{D}^2 g \leqq \tfrac{1}{2}\ \{Z^+ g + Z^- g\}\ \leqq\ \overline{Z} g$$

Hence E_k is contained in the set $\overline{E}_k$ of Theorem 3.

Since $m_e(\overline{E}_k) \geqq m_e(E_k)$, the result follows at once.

COROLLARY. *Let $g(x)$ be continuous and convex in an interval (a, c) and let (α, γ) be any interval interior to (a, c). Then $\overline{D}^2 g < \infty$ almost everywhere in (α, γ).*

Proof: Let E be the set of points in (α, γ) where $\overline{D}^2 g = \infty$. Then E is contained in the set E_k of Theorem 4 for every $k > 0$. Thus $m_e(E_k) \geqq m_e(E)$ and

$$\frac{g(c) - g(\gamma)}{c - \gamma} - \frac{g(\alpha) - g(a)}{\alpha - a} \geqq \tfrac{1}{4}\ k m_e(E) \tag{3.6}$$

for every $k > 0$.

For fixed α, γ the inequality (3.6) cannot hold for every $k > 0$ unless $m_e(E) = 0$. Hence the set of points at which $\overline{D}^2 g = \infty$ has measure zero.

4. DEFINITION OF THE INTEGRAL

Let $f(x)$ be defined in an interval (a, c). The functions $M(x)$ and $m(x)$ are called major *and* minor *functions, respectively, of $f(x)$ in (a, c) if*

(4.1) *$M(x)$ and $m(x)$ are continuous and smooth*;

(4.2) $M(a) = M(c) = m(a) = m(c) = 0$;

$\underline{D}^2 M \geqq f(x)$, $\underline{D}^2 M > -\infty$,

(4.3) $\overline{D}^2 m \leqq f(x)$, $\overline{D}^2 m < +\infty$,

for all x in (a, c) with the possible exception of a denumerable set.

THEOREM 5. *If $M(x)$ is any major function and $m(x)$ any minor function of $f(x)$ in (a, c), then $M(x) - m(x)$ is convex in (a, c) and $M(x) \leqq m(x)$ for every x in (a, c).*

Proof: With the possible exception of a denumerable set of values of x in (a, c), the difference $\underline{D}^2 M - \overline{D}^2 m$ is defined and, by (4.3), is non-negative. Hence $\overline{D}^2(M - m) \geqq \underline{D}^2 M - \overline{D}^2 m \geqq 0$ for all values of x in (a, c) with the possible exception of a denumerable set. By Theorem 1, $M(x) - m(x)$ is convex.

From (2.1), with $g(x) = M(x) - m(x)$, $x_1 = a$, $x_2 = x$, $x_3 = c$, it follows that

$$M(x) - m(x) \leqq \frac{c - x}{c - a}\{M(a) - m(a)\} + \frac{x - a}{c - a}\{M(c) - m(c)\}.$$

But, by (4.2), $M(a) = M(c) = m(a) = m(c) = 0$, and hence $M(x) - m(x) \leqq 0$.

DEFINITION 1 *A function $f(x)$ defined in an interval (a, c) is said to be integrable over (a, b, c), where $a < b < c$, if for every $\epsilon > 0$ there is a major function $M(x)$ and a minor function $m(x)$ such that*

$$0 \leqq m_1(b) - M_1(b) < \epsilon$$

THEOREM 6. *If $f(x)$ is integrable over (a, b, c) there exists a function $F(x)$, defined for all x in (a, c), which is the least upper bound of all major functions $M(x)$, and the greatest lower bound of all minor functions $m(x)$, of $f(x)$ in (a, c).*

Proof: Let $M_0(x)$ be any major function of $f(x)$ in (a, c). By Theorem 5, $M_0(x) \leqq m(x)$ for every minor function $m(x)$. Hence, if $F(x)$ is defined to be the greatest lower bound of all minor functions, then $M_0(x) \leqq F(x)$. But this inequality holds for any major function $M_0(x)$ and therefore $F(x)$ is *an* upper bound of all major functions.

By Definition 1, there exist major and minor functions $M_1(x)$, $m_1(x)$ respectively, for which $0 \leqq m_1(b) - M_1(b) < \epsilon$. Since $M_1(x) - m_1(x)$ is convex, it follows from (2.1), with $g(x) = M_1(x) - m_1(x)$ and x_1, x_2, x_3 replaced by x, b, c respectively, that

$$M_1(b) - m_1(b) \leqq \frac{c - b}{c - x}\{M_1(x) - m_1(x)\} + \frac{b - x}{c - x}\{M_1(c) - m_1(c)\}.$$

But $M_1(c) = m_1(c) = 0$, $M_1(b) \leqq m_1(b)$, and $M_1(x) \leqq m_1(x)$. Hence if $a \leqq x < b$, then

$$0 \leqq m_1(x) - M_1(x) \leqq \frac{c - x}{c - b}\{m_1(b) - M_1(b)\} < \frac{c - a}{c - b}\epsilon = \epsilon_1,$$

and

$$0 \leqq F(x) - M_1(x) \leqq m_1(x) - M_1(x) < \epsilon_1.$$

Similarly, if $b < x \leqq c$, then

$$0 \leqq m_1(x) - M_1(x) \leqq \frac{x - a}{b - a}\{m_1(b) - M_1(b)\} < \frac{c - a}{b - a}\epsilon = \epsilon_2,$$

and

$$0 \leqq F(x) - M_1(x) \leqq m_1(x) - M_1(x) < \epsilon_2.$$

Clearly, $0 \leqq F(b)-M_1(b) \leqq m_1(b)-M_1(b) < \epsilon$. Hence $F(x)$ is the least upper bound of all major functions as well as the greatest lower bound of all minor functions.

DEFINITION 2. *The symbol* $\int_{a,\,b,\,c} f(x)dx$ *denotes the integral of* $f(x)$ *over* (a, b, c). *Its value is* $-F(b)$, *where* $F(x)$ *is the function of Theorem* 6 *Thus,*

$$\int_{a,\,b,\,c} f(x)dx = -F(b).$$

5. ELEMENTARY PROPERTIES OF THE INTEGRAL

The following results are stated without proof. Most of them may be proved by obvious extensions of the methods used in the development of the properties of the ordinary Perron integral.

THEOREM 7. *If* $f(x)$ *is integrable over* (a, b, c), *it is integrable over* (a, x, c) *for every* x *in* (a, c). *If* $F(x)$ *is the function of Theorem* 6, *then*

$$\int_{a,\,x,\,c} f(x)dx = -F(x),$$

for every x *in* (a, c). *Furthermore,* $F(x)$ *is continuous in the interval* (a, c) *and* $M(x)-F(x)$ *is convex for every major function* $M(x)$ *of* $f(x)$ *in* (a, c).

THEOREM 8. *If* $f(x)$ *is integrable over* (a, b, c), *it is integrable over* (α, β, γ) *where* $a \leqq \alpha \leqq \beta \leqq \gamma \leqq c$.

THEOREM 9. *If* $f_1(x)$ *and* $f_2(x)$ *are both integrable over* (a, b, c) *and* $f_1(x) \geqq f_2(x)$ *in* (a, c), *then*

$$\int_{a,\,b,\,c} f_1(x)dx \geqq \int_{a,\,b,\,c} f_2(x)dx.$$

THEOREM 10. *If* $f(x)$ *is integrable over* (a, b, c), *then* $kf(x)$, *where* k *is any constant, is integrable over* (a, b, c) *and*

$$\int_{a,\,b,\,c} kf(x)dx = k\int_{a,\,b,\,c} f(x)dx.$$

THEOREM 11. *If* $f_1(x)$ *and* $f_2(x)$ *are both integrable over* (a, b, c), *then* $f_1(x)+f_2(x)$ *and* $f_1(x)-f_2(x)$ *are integrable over* (a, b, c) *and*

$$\int_{a,\,b,\,c}\{f_1(x) \pm f_2(x)\}dx = \int_{a,\,b,\,c} f_1(x)dx \pm \int_{a,\,b,\,c} f_2(x)dx.$$

THEOREM 12. *If* $f(x)$ *and* $|f(x)|$ *are both integrable over* (a, b, c), *then*

$$\left|\int_{a,\,b,\,c} f(x)dx\right| \leqq \int_{a,\,b,\,c} |f(x)|\,dx.$$

THEOREM 13. *Suppose that the function* $G(x)$ *is continuous and smooth in* (a, c) *and has a finite generalized second derivative* D^2G *for all* x *in* (a, c) *with the possible exception of a denumerable set. Let* $G(x) = D^2G$ *where* D^2G *is finite and* $G(x) = 0$ *elsewhere. Then* $G(x)$ *is integrable over* (a, b, c) *and*

$$\int_{a,\,b,\,c} G(x)dx = \frac{c-b}{c-a}\,G(a) + \frac{b-a}{c-a}\,G(c) - G(b).$$

6. FURTHER PROPERTIES OF THE INTEGRAL

THEOREM 14. *If $f(x)$ is integrable over (a, b, c) and*

$$a \leqq x_1 < x_2 < x_3 \leqq c, \textit{ then}$$

$$\int_{a,\,x_2,\,c} f(x)dx = \int_{x_1,\,x_2,\,x_3} f(x)dx$$

$$+ \frac{x_3-x_2}{x_3-x_1}\int_{a,\,x_1,\,c} f(x)dx + \frac{x_2-x_1}{x_3-x_1}\int_{a,\,x_3,\,c} f(x)dx.$$

Proof: Each of the integrals exists by Theorems 7 or 8. It remains to establish the formula. If $M_1(x)$ and $m_1(x)$ are major and minor functions, respectively, of $f(x)$ in (a, c), it follows that

$$T_1(x) \equiv M_1(x) - \frac{x_3-x}{x_3-x_1}\,M_1(x_1) - \frac{x-x_1}{x_3-x_1}\,M_1(x_3),$$

$$t_1(x) \equiv m_1(x) - \frac{x_3-x}{x_3-x_1}\,m_1(x_1) - \frac{x-x_1}{x_3-x_1}\,m_1(x_3)$$

are major and minor functions, respectively, of $f(x)$ in (x_1, x_3). Hence

$$-t_1(x_2) \leqq \int_{x_1,\,x_2,\,x_3} f(x)dx \leqq -T_1(x_2),$$

$$-\frac{x_3-x_2}{x_3-x_1}\,m_1(x_1) \leqq \frac{x_3-x_2}{x_3-x_1}\int_{a,\,x_1,\,c} f(x)dx \leqq -\frac{x_3-x_2}{x_3-x_1}\,M_1(x_1),$$

and

$$-\frac{x_2-x_1}{x_3-x_1}\,m_1(x_3) \leqq \frac{x_2-x_1}{x_3-x_1}\int_{a,\,x_3,\,c} f(x)dx \leqq -\frac{x_2-x_1}{x_3-x_1}\,M_1(x_3).$$

The addition of the three sets of inequalities gives

$$-m_1(x_2)$$

$$\leqq \int_{x_1,\,x_2,\,x_3} f(x)dx + \frac{x_3-x_2}{x_3-x_1}\int_{a,\,x_1,\,c} f(x)dx + \frac{x_2-x_1}{x_3-x_1}\int_{a,\,x_3,\,c} f(x)dx$$

$$\leqq -M_1(x_2).$$

But

$$-m_1(x_2) \leqq \int_{a,\,x_2,\,c} f(x)dx \leqq -M_1(x_2)$$

and $0 \leqq m_1(x_2) - M_1(x_2) < \epsilon$. It follows that the left- and right-hand members of the formula of the theorem lie in the same interval of length less than ϵ. Since ϵ is arbitrary, this establishes the formula.

THEOREM 15. *If $f(x)$ is integrable over (a, b, c) and $f(x) \geqq 0$ in (a, c) then*

$$F(x) = -\int_{a,\,x,\,c} f(x)dx$$

is convex.

Proof: By Theorem 14,

$$\frac{x_3-x_2}{x_3-x_1}F(x_1)+\frac{x_2-x_1}{x_3-x_1}F(x_3)-F(x_2)$$

$$=-\frac{x_3-x_2}{x_3-x_1}\int_{a,\,x_1,\,c}f(x)dx-\frac{x_2-x_1}{x_3-x_1}\int_{a,\,x_3,\,c}f(x)dx+\int_{a,\,x_2,\,c}f(x)dx$$

$$=\int_{x_1,\,x_2,\,x_3}f(x)dx$$

But, if $f(x) \geqq 0$, then Theorem 9 with $f_2(x) \equiv 0$ shows that

$$\int_{x_1,\,x_2,\,x_3}f(x)dx \geqq 0.$$

Hence $F(x)$ is convex.

THEOREM 16 *If $f(x)$ and $|f(x)|$ are both integrable over (a, b, c), then*

$$\int_{a,\,x,\,c}f(x)dx = F_2(x) - F_1(x),$$

where $F_1(x)$ and $F_2(x)$ are convex in (a, c).

Proof: Let $f_1(x) = f(x)$, for all values of x such that $f(x) \geqq 0$, and let $f_1(x) = 0$ elsewhere; let $f_2(x) = f_1(x) - f(x)$. Then $f_1(x) \geqq 0$, $f_2(x) \geqq 0$ for all x in (a, c). Also

$$f_1(x) = \tfrac{1}{2}\left\{|f(x)| + f(x)\right\},\quad f_2(x) = \tfrac{1}{2}\left\{|f(x)| - f(x)\right\}$$

so that $f_1(x)$ and $f_2(x)$ are both integrable by Theorems 10 and 11. Therefore

$$\int_{a,\,x,\,c}f(x)dx = \int_{a,\,x,\,c}f_1(x)dx - \int_{a,\,x,\,c}f_2(x)dx = -F_1(x) + F_2(x),$$

where $F_1(x)$ and $F_2(x)$ are convex by Theorem 15.

7. DIFFERENTIABILITY OF THE INTEGRAL

The proof of the differentiability of the integral parallels that given by McShane for the ordinary Perron integral.

THEOREM 17. *Let $f(x)$ be integrable over (a, b, c) and*

$$F(x) = -\int_{a,\,x,\,c}f(x)dx.$$

Let (α, γ) be an interval interior to (a, c). Then, for almost all x in (α, γ), the function $F(x)$ has a finite generalized second derivative equal to $f(x)$.

Proof: Let ϵ and k be arbitrary positive numbers. Let $M_1(x)$ and $m_1(x)$ be major and minor functions, respectively, of $f(x)$ in (a, c) such that $0 \leqq m_1(b) - M_1(b) < \epsilon_1$, where $\alpha \leqq b \leqq \gamma$ and

$$\epsilon_1 = \frac{k\epsilon}{4\left\{\dfrac{\gamma-a}{(c-\gamma)(b-a)}+\dfrac{c-\alpha}{(\alpha-a)(c-b)}\right\}}. \tag{7.1}$$

By Theorem 7, $M_1(x) - F(x)$ is convex in (a, c) and hence, by the corollary to Theorem 4, $\overline{D}^2(M_1 - F) < \infty$ except on a set of measure zero. Then the inequality

$$\underline{D}^2F = \underline{D}^2\{M_1-(M_1-F)\} \geqq \underline{D}^2M_1-\overline{D}^2(M_1-F)$$

and (4.3) imply that $\underline{D}^2F > -\infty$ almost everywhere.

Let E_k be the set of points in (α, γ) for which $\overline{D}^2(M_1-F) \geqq k$. From Theorem 4 it follows that

$$\frac{\{M_1(c)-F(c)\}-\{M_1(\gamma)-F(\gamma)\}}{c-\gamma} - \frac{\{M_1(\alpha)-F(\alpha)\}-\{M_1(a)-F(a)\}}{\alpha-a}$$
$$\geqq \tfrac{1}{4}\, km_e(E_k).$$

Since $M_1(c) = M_1(a) = F(c) = F(a) = 0$, this may be written

$$\frac{F(\gamma)-M_1(\gamma)}{c-\gamma} + \frac{F(\alpha)-M_1(\alpha)}{\alpha-a} \geqq \tfrac{1}{4}\, km_e(E_k).$$

But $m_1(\gamma) \geqq F(\gamma)$, $m_1(\alpha) \geqq F(\alpha)$ and therefore

$$\frac{m_1(\gamma)-M_1(\gamma)}{c-\gamma} + \frac{m_1(\alpha)-M_1(\alpha)}{\alpha-a} \geqq \tfrac{1}{4}\, km_e(E_k). \tag{7.2}$$

As in the proof of Theorem 6,

$$m_1(\gamma)-M_1(\gamma) \leqq \frac{\gamma-a}{b-a}\epsilon_1, \quad m_1(\alpha)-M_1(\alpha) \leqq \frac{c-\alpha}{c-b}\epsilon_1.$$

Hence the left-hand member of (7.2) is less than

$$\left\{\frac{\gamma-a}{(c-\gamma)(b-a)} + \frac{c-\alpha}{(\alpha-a)(c-b)}\right\}\epsilon_1$$

which equals $\frac{1}{4}k\epsilon$ by (7.1). Thus $m_e(E_k) < \epsilon$.

Let E_0 be the set of points in (α, γ) for which one or more of the inequalities (4.3) fails to hold. If x is not in E_0+E_k, then

$$\underline{D}^2F \geqq \underline{D}^2M_1-\overline{D}^2(M_1-F) \geqq f(x)-k.$$

Since $m_e(E_0+E_k) \leqq m_e(E_0)+m_e(E_k) < \epsilon$, and ϵ is arbitrary, $m_e(E_0+E_k) = 0$. Hence

$$\underline{D}^2F \geqq f(x)-k$$

except on a set of measure zero.

In particular, if $k = 1/n$, where n is a positive integer, then

$$\underline{D}^2F \geqq f(x)-\frac{1}{n} \tag{7.3}$$

except on a set E_n of measure zero. Except on the set $E = E_1+E_2+\cdots$, which also has measure zero, the inequality (7.3) is valid for every positive integer. Hence

$$\underline{D}^2F \geqq f(x), \quad \underline{D}^2F > -\infty \tag{7.4}$$

almost everywhere.

By Theorem 10 the function $-f(x)$ is integrable and $F(x)$ is its integral. Therefore

$$\underline{D}^2(-F) \geqq -f(x),\ \underline{D}^2(-F) > -\infty$$

almost everywhere. But $\underline{D}^2(-F) = -\overline{D}^2F$ and hence

(7.5) $$\overline{D}^2F \leqq f(x),\ \overline{D}^2F < +\infty$$

almost everywhere

Since $\underline{D}^2F \leqq \overline{D}^2F$ for every x in (a, c) it follows from (7.4) and (7.5) that

$$-\infty < \underline{D}^2F \leqq \overline{D}^2F \leqq f(x) \leqq \underline{D}^2F \leqq \overline{D}^2F < +\infty$$

almost everywhere. This in turn implies that

$$D^2F = \overline{D}^2F = \underline{D}^2F = f(x),$$
$$-\infty < D^2F < +\infty$$

for almost all x in (a, c).

8. The Generality of the P^2-Integral

Let $G(x) = x^4 \cos (1/x^2)$ and let

$$G(x) = \begin{cases} D^2G = G''(x) \text{ if } x \neq 0, \\ 0 \text{ if } x = 0. \end{cases}$$

The function $G(x)$ is *not* Perron integrable over any interval which includes $x = 0$ (Burkill). Since $G(x)$ satisfies the conditions of Theorem 13, however, the function $G(x)$ is P^2-integrable over any finite interval.

9. Conclusion

In this paper it has only been possible to consider the simpler properties of the P^2-integral. Other properties and applications of the integral will be discussed in another paper. It is not difficult to show, for example, that the conditions $\underline{D}^2M \geqq f(x)$, $\overline{D}^2m \leqq f(x)$ in the definition of major and minor functions need only be satisfied for almost all x in the interval.

Again, the function $G(x)$ of §8, while not Perron integrable, is Cesàro-Perron integrable (Burkill). Whether the P^2-integral includes the Cesàro-Perron integral or not is still an open question.

Finally, the problems of integration by substitution and by parts for the P^2-integral have still to be investigated.

UNIVERSITY OF BRITISH COLUMBIA.

REFERENCES

(1) Burkill, J. C. The Cesàro-Perron Integral Proc London Math Soc, 2, 34· 314-22 1932.

(2) Hardy, G H, Littlewood, J. E., and Polya, G Inequalities Cambridge 1934

(3) ———— and Rogosinski, W W. Fourier Series Cambridge 1944.

(4) McShane, E. J. Integration. Princeton. 1944

(5) Saks, S Theory of the Integral Warsaw 1937

(6) Titchmarsh, E C The Theory of Functions Oxford 1932

A GROUP THEORETIC CHARACTERIZATION OF THE GENERAL PROJECTIVE COLLINEATION GROUP

By N. S MENDELSOHN

Presented by G. DE B. ROBINSON, F R.S.C.

INTRODUCTION

IT is our purpose to give a set of axioms sufficient to characterize the projective collineation group From the properties of the group so obtained a projective geometry is constructed, thus proving that projective geometry is completely characterized by its projective collineation group. The automorphisms of the group are studied, the study yielding an elementary geometric proof of the theorem that all automorphisms of the projective collineation group form the group of all collineations and correlations (projective or not) of the related geometry.

The paper is divided into three parts In the first part, certain properties of the projective group will be developed. These properties will be the ones upon which our axiomatic development in part II will be based. Part III will consist of a discussion of the group automorphisms Co-ordinate free methods are used throughout in part III. For the most part we consider the two dimensional case but all the results may be generalized.

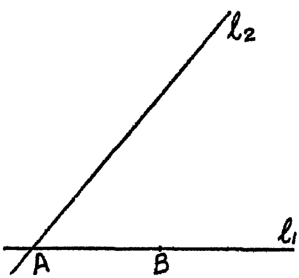
PART I· PROPERTIES OF PROJECTIVE COLLINEATIONS

In terms of co-ordinates the projective collineation group consists of all transformations $\rho x'_i = \sum_{j=1}^{n} a_{ij}x_j$ $(i = 1, 2, \ldots n)$ where det $(a_{ij}) \neq 0$, a_{ij} and ρ being elements from a fixed field F, and the dimensionality of the geometry being $n - 1$. The transformation $\rho x_i' = \sum_{j=1}^{n} a_{ij}x_j$ is characterized by the matrix $A = (a_{ij})$ provided we identify the matrices A and kA. A change of co-ordinates with matrix P replaces A by $P^{-1}AP$. In classifying collineations the transformations A and $P^{-1}AP$ must, therefore, be considered in the same class. From classical matrix theory, a necessary and sufficient condition that $B = P^{-1}AP$ is that the matrices $\lambda I - A$ and $\lambda I - B$ have the same invariant factors, or equivalently, the same elementary divisors. In the case of a co-ordinate field which is algebraically closed the Segre characteristic suffices to classify all collineations. In the two-dimensional case (with algebraically closed field) the classification is as follows.

TYPE 1. Characteristic [1, 1, 1]. This is the general case. There are three fixed points and three fixed lines. If the fixed points be taken as the triangle of reference,

the matrix takes the form $\begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix}$, where a_{11}, a_{22}, a_{33} are all different and different from 0.

TYPE 2. Characteristic [2, 1]. In this case there are two fixed points and two fixed lines arranged as in the diagram. If the triangle of reference has its vertices at A and B and its third vertex any other point on l_2, the matrix takes the form



$$\begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & a_{32} & a_{22} \end{pmatrix},$$

where $a_{11}a_{22}a_{32} \neq 0$ and $a_{11} \neq a_{22}$.

TYPE 3. Characteristic [3]. Here there is one fixed point and one fixed line passing through it. If the fixed point be (0, 0, 1) and the fixed line be $x_1 = 0$ the transformation takes the form $\begin{pmatrix} 1, & 0, & 0 \\ a_{21}, & 1, & 0 \\ a_{31}, & a_{32}, & 1 \end{pmatrix}$, where a_{21}, a_{32} are both different from 0.

TYPE 4. Characteristic [(1, 1), 1]. *The homology.* In this important case there is a line of fixed points called the axis and a point of fixed lines called the centre, the centre not lying on the axis. If the centre is at (0, 0, 1) and the axis is $x_3 = 0$, the homology takes the form

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & a_{33} \end{pmatrix} \text{ where } a_{33} \neq 1, 0.$$

If $a_{33} = -1$ the transformation is called a harmonic homology. A homology is completely determined geometrically by its centre, axis, and a pair of corresponding points (not a double point.)

TYPE 5. Characteristic [(2, 1)]. *The planar elation.* This type differs from the homology in that the centre lies on the axis. It is completely characterized by its centre axis and a pair of corresponding points. With centre (0, 0, 1) and axis $x_1 = 0$, the elation is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & a_{32} & 1 \end{pmatrix} \text{ where } a_{32} \neq 0.$$

TYPE 6. Characteristic [(1, 1, 1)]. The identity $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

In the case where the co-ordinate field is not algebraically closed there are further classes, but in all cases (except the trivial field consisting of 0 and 1 alone) these six exist. The whole group G may be generated by transformations of type 4 (homologies). The group generated by transformations of type 5 is a sub-group of G (in general), called the unimodular projective group.

IMPORTANT SUB-GROUPS OF G

In this section, a study of those sub-groups which are to be the basis of our axiomatic system will be made. We confine ourselves to the planar case.

Let P be a point of a projective plane.

Associated with P is the sub-group $\mathfrak{P}$ of G which consists of all transformations in G which keep every line through P fixed. $\mathfrak{P}$ is generated by all homologies with centres at P and consists of all homologies and elations with centre at P. Dually, let L be a line of the projective plane. $\mathfrak{L}$ is defined to be the group of all collineations of G which keep every point on L fixed.

We state a number of properties of the groups $\{\mathfrak{P}\}$ and $\{\mathfrak{L}\}$.

Let $N(\mathfrak{P})$ represent the normalizer of $\mathfrak{P}$ in G. $N(\mathfrak{P})$ consists of all projective collineations which keep the point P fixed. Similarly, $N(\mathfrak{L})$ consists of all projective collineations which keep the line L fixed.

(1.1) If P_1 and P_2 are two points, the intersection of their groups $\mathfrak{P}_1 \wedge \mathfrak{P}_2 = 1$. This follows from the fact that a projective transformation which keeps four points fixed, no three of which are collinear, is the identity.

(1.2) If L_1 and L_2 are two lines then $\mathfrak{L}_1 \wedge \mathfrak{L}_2 = 1$ (dual of 1.1).

(1.3) If P is a point $NN(\mathfrak{P}) = N(\mathfrak{P})$.

(1.4) If L is a line $NN(\mathfrak{L}) = N(\mathfrak{L})$.

(1.5) If P_1 and P_2 are two distinct points it is impossible for $N(\mathfrak{P}_1) \leq N(\mathfrak{P}_1)$. For if α is in $N(\mathfrak{P}_1)$ such that $\alpha: P_2 \rightarrow P_3$, $P_2 \neq P_3$, then α is not in $N(\mathfrak{P}_2)$.

(1.6) If L_1 and L_2 are two distinct lines it is impossible for

$$N(\mathfrak{L}_1) \leq N(\mathfrak{L}_2).$$

(1.7) If, and only if, P_1, P_2, P_3 are three collinear points $\mathfrak{P}_1 \leq (\mathfrak{P}_2, \mathfrak{P}_3)$. (The notation $(\mathfrak{P}_2, \mathfrak{P}_3)$ represents the union of, i.e. the group generated by, $\mathfrak{P}_2$ and $\mathfrak{P}_3$.)

(1.8) The lines L_1, L_2, L_3 are concurrent, if and only if $\mathfrak{L}_1 \leq (\mathfrak{L}_2, \mathfrak{L}_3)$.

(1.9) A point P lies on a line L if and only if $\mathfrak{P} \leq N(\mathfrak{L})$ and this is equivalent to $\mathfrak{L} \leq N(\mathfrak{P})$.

(1.10) If P_1, P_2, P_3, are three distinct points, collinear on L, then $N(\mathfrak{P}_1) \wedge N(\mathfrak{P}_2) \wedge N(\mathfrak{P}_3) = \mathfrak{L}$.

(1.11) If L_1, L_2, L_3 are three distinct lines, concurrent on P, then $N(\mathfrak{L}_1) \wedge N(\mathfrak{L}_2) \wedge N(\mathfrak{L}_3) = \mathfrak{P}$.

(1.12) If L is the join of the points P_1 and P_2 then $(\mathfrak{P}_1, \mathfrak{P}_2) = N(\mathfrak{L})$.

(1.13) If the lines L_1, L_2 meet in the point P then $(\mathfrak{L}_1, \mathfrak{L}_2) = N(\mathfrak{P})$.

(1.14) If P_1 and P_2 are two points $\mathfrak{P}_2 = g^{-1}\mathfrak{P}_1 g$ where g is of finite order.

(1.15) If L_1 and L_2 are two lines $\mathfrak{L}_2 = g^{-1}\mathfrak{L}_1 g$ where g is of finite order.

(1.16) If P_1, P_2, P_3 are three non-collinear points $(\mathfrak{P}_1, \mathfrak{P}_2, \mathfrak{P}_3) = G$.

(1.17) If L_1, L_2, L_3 are three non-concurrent lines $(\mathfrak{L}_1, \mathfrak{L}_2, \mathfrak{L}_3) = G$.

(1.18) Let P be any point and L be any line. Let G be separated into right cosets relative to $N(\mathfrak{L})$; $G = N(\mathfrak{L}) + \Sigma N(\mathfrak{L})g$. Then, either $\mathfrak{P} \leq N(\mathfrak{L})$ or each coset $N(\mathfrak{L})g$ which is such that $\mathfrak{P} \not\leq g^{-1}N(\mathfrak{L})g$ has the property that $N(\mathfrak{L})g \wedge \mathfrak{P}$ is a non-null set.

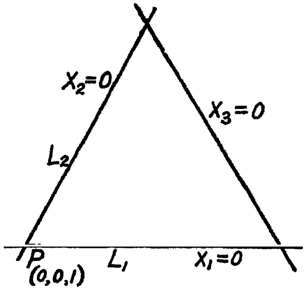
Proofs of (1.7) to (1.18) are outlined below.

Proof of (1.9): Let the point P lie on the line L. Let a be in the group $\mathfrak{P}$. Since a keeps every line through P fixed, a is in $N(\mathfrak{L})$. Dually, if L passes through P, $L \leq N(\mathfrak{P})$. If P is not on L, let a be a homology with centre P but with axis different from L. Then a is in $\mathfrak{P}$ but a is not in $N(\mathfrak{L})$. Thus $\mathfrak{P} \not\leq N(\mathfrak{L})$.

Proof of (1.10): Since P_1, P_2, P_3 all lie on L, Theorem 1.9 implies $\mathfrak{L} \leq N(\mathfrak{P}_1)$, $\mathfrak{L} \leq N(\mathfrak{P}_2)$, $\mathfrak{L} \leq N(\mathfrak{P}_3)$. Hence, $\mathfrak{L} \leq N(\mathfrak{P}_1) \wedge N(\mathfrak{P}_2) \wedge N(\mathfrak{P}_3)$. Let a be an element of $N(\mathfrak{P}_1) \wedge N(\mathfrak{P}_2) \wedge N(\mathfrak{P}_3)$. Then a keeps P_1, P_2, P_3 fixed. By the fundamental theorem of projective geometry a keeps all points on the join L of P_1, P_2, P_3 fixed, i.e. a is in $\mathfrak{L}$. Thus $\mathfrak{L} = N(\mathfrak{P}_1) \wedge N(\mathfrak{P}_2) \wedge N(\mathfrak{P}_3)$. This theorem does not hold in projective geometries which do not satisfy Pappus' Theorem. Theorem 1.11 is the dual of Theorem 1.10 and is, therefore, also true.

Proof of (1.13): Since L_1 and L_2 both pass through P, $\mathfrak{L}_1 \leq N(\mathfrak{P})$ and $\mathfrak{L}_2 \leq N(\mathfrak{P})$ by (1.9).

Thus $(\mathfrak{L}_1, \mathfrak{L}_2) \leq N(\mathfrak{P})$. We now show that $(\mathfrak{L}_1, \mathfrak{L}_2) \geq N(\mathfrak{P})$. Take L_1 to be the line $x_1 = 0$ and L_2 to be the line $x_2 = 0$. Then P has co-ordinates (0, 0, 1).



The groups $\mathfrak{L}_1$ and $\mathfrak{L}_2$ contain the following elements

1. Homology with $x_1 = 0$ as axis

$$\begin{pmatrix} \lambda & 0 & 0 \\ a_{21} & 1 & 0 \\ a_{31} & 0 & 1 \end{pmatrix},$$ where λ is any element of the co-ordinate field F different from 0, 1 and a_{21}, a_{31} are arbitrary elements in F. We denote this type of element L_1^λ. Use will only be made of the cases $a_{21} = a_{31} = 0$.

2. Elation with $x_1 = 0$ as axis, (0, 1, 0) as centre

$$\begin{pmatrix} 1 & 0 & 0 \\ \lambda & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$ Denote this type by ${L_1}^{(\lambda)}{}_2$.

3. Elation with $x_1 = 0$ as axis, (0, 0, 1) as centre

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \lambda & 0 & 1 \end{pmatrix} = {L_1}^{(\lambda)}{}_3$$

Similarly for L_2 the types

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{pmatrix} = L_2^\lambda; \begin{pmatrix} 1 & \lambda & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = {L_2}^{(\lambda)}{}_1; \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda & 1 \end{pmatrix} = {L_2}^{(\lambda)}{}_3$$ are obtained.

The most general element of $N(\mathfrak{P})$ is of the form

$$\begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 1 \end{pmatrix} = K.$$

The elements a_{11}, a_{12} are not both 0. Suppose that $a_{11} \neq 0$.

Then $$K{L_2}^{\left(-\frac{a_{12}}{a_{11}}\right)}{}_1 = \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & \bar{a}_{22} & 0 \\ a_{31} & \bar{a}_{32} & 1 \end{pmatrix}.$$ Multiplying on the right by $L_1^{\frac{1}{a_{11}}}$, we obtain $$K{L_2}^{\left(-\frac{a_{12}}{a_{11}}\right)}{}_1 L_1^{\frac{1}{a_{11}}} = \begin{pmatrix} 1 & 0 & 0 \\ \bar{a}_{21} & \bar{a}_{22} & 0 \\ \bar{a}_{31} & \bar{a}_{32} & 1 \end{pmatrix}.$$ Hence ${L_1}^{(-\bar{a}_{21})}{}_2$

$KL_2{}^{\left(-\frac{a_{12}}{a_{11}}\right)_1} L_1{}^{\frac{1}{a_{11}}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \bar{a}_{22} & 0 \\ \bar{a}_{31} & \bar{a}_{32} & 1 \end{pmatrix}$. Continuing $L_1{}^{(-\bar{a}_{31})_3}$ $L_1{}^{(-\bar{a}_{21})}$

$KL_2{}^{\left(-\frac{a_{12}}{a_{11}}\right)_1} L_1{}^{\frac{1}{a_{11}}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \bar{a}_{22} & 0 \\ 0 & \bar{a}_{32} & 1 \end{pmatrix}$. Finally, $L_2{}^{(-\bar{a}_{32})_3}$, $L_2{}^{\frac{1}{\bar{a}_{22}}}$ $L_1{}^{(-\bar{a}_{21})_2}$

$L_1{}^{(-a_{31})}$ $KL_2{}^{\left(-\frac{a_{12}}{a_{11}}\right)_1} L_1{}^{\frac{1}{a_{11}}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Therefore, $K =$ $\left[L_2{}^{\left(-\frac{a_{12}}{a_{11}}\right)_1} L_1{}^{\frac{1}{a_{11}}} L_2{}^{(-\bar{a}_{32})_3} L_2{}^{\frac{1}{\bar{a}_{22}}} L_1{}^{(-a_{31})_3} L_1{}^{(-\bar{a}_{21})_2}\right]^{-1}$. If $a_{11} = 0$, then $a_{12} \neq 0$ and hence

$$KL_1{}^{(1)} \cdot = \begin{pmatrix} a_{12} & a_{12} & 0 \\ \bar{a}_{21} & a_{22} & 0 \\ \bar{a}_{31} & a_{32} & 1 \end{pmatrix} \text{ and we may proceed}$$

as in the first case. Thus $(\mathfrak{L}_1, \mathfrak{L}_2) \geq N(\mathfrak{P})$. Combining this with $(\mathfrak{L}_1, \mathfrak{L}_2) \leq N(\mathfrak{P})$, we obtain $(\mathfrak{L}_1, \mathfrak{L}_2) = N(\mathfrak{P})$.

Theorem 1.12 follows in a dual manner.

Proof of (1.8): Let L_1, L_2, L_3 be concurrent at P. From (1.13) $(\mathfrak{L}_1, \mathfrak{L}_2) = N(\mathfrak{P})$. From (1.9) $\mathfrak{L}_3 \leq N(\mathfrak{P})$. Thus $\mathfrak{L}_1 \leq (\mathfrak{L}_2, \mathfrak{L}_3)$. Dually, (1.7) follows.

Proof of (1.17): Since L_1, L_2, L_3, are not concurrent, we may choose as co-ordinate axes $L_1: x_1 = 0$, $L_2: x_2 = 0$, $L_3: x_3 = 0$. The group $\mathfrak{L}_1$ contains each of the transformations (with $\lambda \neq 0$);

$$\begin{pmatrix} \lambda & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{— homology with } L_1 \text{ as axis, centre } (1, 0, 0);$$

$$\begin{pmatrix} 1 & 0 & 0 \\ \lambda & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{— elation with } L_1 \text{ as axis, centre } (0, 1, 0);$$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \lambda & 0 & 1 \end{pmatrix} \text{— elation with } L_1 \text{ as axis, centre } (0, 0, 1).$$

Similarly $\mathfrak{L}_2$ contains each of the matrices

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & \lambda & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda & 1 \end{pmatrix}; \text{ and}$$

$\mathfrak{L}_3$ contains each of the matrices

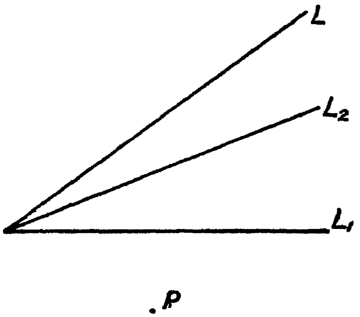
$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \lambda \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & \lambda \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \lambda \\ 0 & 0 & 1 \end{pmatrix}$$

Multiplying any matrix by a suitable matrix of the above nine types affects one of the following transformations

1. The elements of a row (column) are multiplied by the same non-zero factor

2. To the elements of one row (column) are added the same multiple of the corresponding elements of another row (column). It is clear that these transformations are sufficient to reduce any non-singular matrix to the identity Hence (1.17) and dually (1.16) follow

Proof of (1.18): If P is on L then $\mathfrak{P} \leq N(\mathfrak{L})$. If P is not on L and $\mathfrak{P} \not\leq g^{-1}N(\mathfrak{L})g$, let $g^{-1}\mathfrak{L}g = \mathfrak{L}_1$ Then $g^{-1}N(\mathfrak{L})g = N(\mathfrak{L}_1)$. (A proof of this property will be given in part II where much use will be made of it) Thus $\mathfrak{P} \not\leq N(\mathfrak{L}_1)$, i.e., P is not on L_1. Let L_2 be any line through the intersection of L and L_1 which does not pass through P. Take a homology t with centre at P and axis at L_2 such that t $L \rightarrow L_1$. Then t is in $\mathfrak{P}$ and $t^{-1}\mathfrak{L}t = \mathfrak{L}_1$. Thus $t^{-1}\mathfrak{L}t = g^{-1}\mathfrak{L}g$ or $gt^{-1}\mathfrak{L}tg^{-1} = \mathfrak{L}$. Therefore, tg^{-1} is in $N(\mathfrak{L})$ or t is in $N(\mathfrak{L})g$. Hence t is in $\mathfrak{P} \wedge N(\mathfrak{L})g$ The proof breaks down for the seven-point geometry PG (2,2), but a direct verification may be made in this case.



Proof of (1.15): The theorem is obvious in the case G is of finite order. If G is of infinite order and the co-ordinate field is not of characteristic 2, take a harmonic homology t, such that t: $L_1 \rightarrow L_2$, $L_2 \rightarrow L_1$. Then $t^2 = 1$ and $t^{-1}\mathfrak{L}_1t = \mathfrak{L}_2$. If the co-ordinate field is of infinite order and of characteristic 2, choose the co-ordinate axes so that L_1 is $x_2 = 0$ and L_2 is $x_1 = 0$. Consider the transformation

$$t = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Making use of the fact that $1 + 1 = 0$ we obtain $t^3 = 1$ and $t^{-1}\mathfrak{L}_1t = \mathfrak{L}_2$. Dually (1.14) is true.

PART II CHARACTERIZATION OF G

In this part we give an abstract characterization of the plane projective collineation group G. Extension to n-dimensions can be carried out in a direct manner. We now do not start out with a geometry Thus it is no longer necessary nor indeed possible to distinguish between a point P and its group $\mathfrak{P}$. Only the latter exists at first. From here on we will use only the symbol P. Similar remarks apply to lines which will usually be denoted by L. We introduce the notion of strongly conjugate groups. Two groups R and S are strongly conjugate in G if there exists an element g in G, such that $R = g^{-1}Sg$ and g is of *finite order.*

AXIOMS

1. Let G be a group generated by two complete sets of strongly conjugate sub-groups $\{P\}$ and $\{L\}$, called points and lines for convenience, such that the only element common to every sub-group of both sets is the identity.

2. If T is a point or line $NN(T) = N(T)$.

3. If $A \leq N(D)$ there exists two other sub-groups B and C, each strongly conjugate to A such that $B \leq N(D)$, $C \leq N(D)$. (Here A is a point or line according as D is a line or point).

4. (*a*) If P and Q are distinct points $(P, Q) = N(L)$ for some line L.
   (*b*) If L and M are distinct lines $(L, M) = N(P)$ for some point P.

DEFINITIONS. A point P is said to be on a line L if $P \leq N(L)$. A line L is said to be on a point P if $L \leq N(P)$. At a later stage we will show that if P is on L then L is on P.

5. If A, B, C are three distinct points (lines) on the line (point) L, then $N(A) \wedge N(B) \wedge N(C) = L$.

6. Let P and L be any point and line respectively. Suppose G be separated into right cosets with respect to $N(L)$; $G = N(L) + \Sigma N(L)g$. Either $P \leq N(L)$ or for each coset $N(L)g$ such that $P \not\leq g^{-1}N(L)g$, the set $P \wedge N(L)g$ is non-null.

At a later point in the discussion axiom 6 will be replaced by a slightly stronger one. Apart from this the above six axioms constitute a complete list of restrictions we place on G. The axioms are consistent since they hold in any finite projective geometry for which every possible case can be verified.

THEOREMS

(2.1) $P \leq N(L)$ if and only if $L \leq N(P)$.

Proof: Let $P \leq N(L)$. By axiom 3 there exist two points Q and R distinct from P such that $Q \leq N(L)$, $R \leq N(L)$. Therefore $N(P) \wedge N(Q) \wedge N(R) = L$ (axiom 5). Hence $L \leq N(P)$. The theorem would also follow from the condition $N(P) \wedge N(Q) \wedge N(R) \geq L$ which is weaker than axiom 5. This is of some consequence since the weaker axiom is satisfied in certain geometries for which Pappus' Theorem does not hold, whereas axiom 5 is not.

(2.2) If $N(A) = N(B)$ than $A = B$ where A and B are both points or lines.

Proof: Let $B = g^{-1}Ag$ (axiom 1). We first show that $N(B) = g^{-1} N(A)g$. Let a be in $N(B)$; $a^{-1}Ba = B$, or $a^{-1}g^{-1}Aga = B = g^{-1}Ag$. Thus $(gag^{-1})^{-1}A(gag^{-1}) = A$; hence gag^{-1} is in $N(A)$ or a is in $g^{-1}N(A)g$. Thus $N(B) \leq g^{-1}N(A)g$. Similarly, $N(A) \leq gN(B)g^{-1}$ or $g^{-1}N(A)g \leq N(B)$. Therefore $N(B) = g^{-1}N(A)g$. If $N(B) = N(A)$ then $g^{-1}N(A)g = N(A)$ or g is in $NN(A)$. By axiom 2, g is in $N(A)$ or $g^{-1}Ag = A = B$

(2.3) If P_1 and P_2 are two points, there is one and only one line L such that P_1 is on L and P_2 is on L. We call L the join of P_1 and P_2.

Proof: By axiom 4(a), $(P_1, P_2) = N(L)$ for some L. Therefore, $P_1 \leq N(L)$, $P_2 \leq N(L)$. Therefore, at least one L exists. If another line L_1 existed such that $P_1 \leq N(L_1)$, $P_2 \leq N(L_1)$ then $(P_1, P_2) \leq N(L_1)$ or $N(L) \leq N(L_1)$. Let $L_1 = g^{-1}Lg$ where g is of finite order (axiom 1). Then $N(L_1) = g^{-1}N(L)g$ (as proved in the previous theorem). Therefore $N(L) \leq g^{-1}N(L)g$. From this we obtain $N(L) \leq g^{-1}N(L)g \leq g^{-2} N(L)g^2 \leq \ldots g^{-n}N(L)g^n \leq \ldots.$ If $g^n = 1$, $g^{-n}N(L)g^n = N(L)$; hence $N(L) = g^{-1}N(L)g = N(L_1)$. By (2.2) $L = L_1$.

COROLLARY 1. If $P \leq N(L)$, $Q \leq N(L)$; $(P, Q) = N(L)$.

COROLLARY 2. The points A, B, C are collinear (i.e., are on the same line if and only if $C \leq (A, B)$.

(2.4) Two lines have one and only one point in common. This is the dual of (2.3) and is proved in the same way.

Theorems 2.2, 2.3, 2.4 and axiom 3, show that the axioms of incidence of projective geometry are satisfied by the sub-groups of G called points and lines.

TRANSFORMATIONS

In this section we build up a theory of projective correspondences.

For this work use is made of the following two lemmas.

(2.5) $p^{-1}(A, B)p = (p^{-1}Ap, p^{-1}Bp)$ for any two sub-groups A and B.

Proof: Let a be in $p^{-1}(A, B)p$. Then $a = p^{-1}(a_1b_1a_2b_2 \ldots a_nb_n)$ $p = (p^{-1}a_1p)(p^{-1}b_1p) \ldots (p^{-1}b_np)$ where a_i is in A, b_i in B. Thus a is in $(p^{-1}Ap, p^{-1}Bp)$. Similarly a in $(p^{-1}Ap, p^{-1}Bp)$ implies a is in p^{-1} $(A, B)p$. Hence $p^{-1}(A, B)p = (p^{-1}Ap, p^{-1}Bp)$.

(2.6) $p^{-1}(A \wedge B)p = p^{-1}Ap \wedge p^{-1}Bp$. The proof is similar to that of (2.5).

DEFINITION. Let A and B be two points of G. The element t of G is said to carry A into B (written t: $A \rightarrow B$) if $t^{-1}At \leq B$. Let $t^{-1}At = C$. Then $C \leq B$. But $B = g^{-1}Cg$ where g is of finite order (axiom 1). Hence $C \leq g^{-1}Cg$. The proof used in (2.3) shows that $C = g^{-1}Cg = B$. Hence if t: $A \rightarrow B$, $t^{-1}At = B$.

(2.7) If t: $A \rightarrow B$ and q: $B \rightarrow C$, then tq: $A \rightarrow C$. For $(tq)^{-1}A(tq) = q^{-1}(t^{-1}At)q = q^{-1}Bq = C$.

(2.8) If t: $A \rightarrow B$ then t^{-1}: $B \rightarrow A$. For $t^{-1}At = B$ is equivalent to $(t^{-1})^{-1}Bt^{-1} = A$.

(2.9) If A is a point or line $N(A)$ consists of all elements t of G which keep A fixed.

For if t: $A \rightarrow A$, $t^{-1}At = A$ or t is in $N(A)$ and conversely.

(2.10) If P is a point and L is a line through P. Then each p in P is such that p: $L \underline{\rightarrow} L$.

Proof: $L \leq N(P)$, therefore $P \leq N(L)$ (by 2.1). If p is in P, p is in $N(L)$ or $p^{-1}Lp = L$, i.e., p: $L \rightarrow L$. The theorem states that every element of P keeps fixed every line through P.

(2.11) If L_1, L_2 and L_3 are three lines through P and if the element t of G keeps L_1, L_2, L_3 fixed, t is in P. For t is in $N(L_i)$ $(i = 1, 2, 3)$. Thus t is in $N(L_1) \wedge N(L_2) \wedge N(L_3) = P$ (axiom 5).

COROLLARY 1. From (2.10) and (2.11) a point consists of all elements of G each of which keeps every line through that point fixed.

COROLLARY 2. If an element t of G keeps three lines through the point P fixed it keeps all lines through P fixed.

Dually, we have:

(2.12) A line L consists of all elements t in G each of which keeps every point on L fixed. If t keeps three points of L fixed, it keeps every point of L fixed.

(2.13) If t keeps two points P_1 and P_2 fixed, it keeps their join fixed.

Proof: Let P_1, P_2 meet in L. Then $(P_1, P_2) = N(L)$. Also $t^{-1}P_1t = P_1$, $t^{-1}P_2t = P_2$. Therefore,

$$t^{-1}N(L)t = t^{-1}(P_1, P_2)t = (t^{-1}P_1t, t^{-1}P_2t) = (P_1, P_2) = N(L).$$

Thus t is in $NN(L)$. By axiom 2, t is in $N(L)$ or t: $L \rightarrow L$. The dual is proved in the same way; hence

(2.14) If t keeps two lines L_1 and L_2 fixed, it keeps their meet fixed.

(2.15) If A, B, C are collinear points and $t: A \rightarrow A', B \rightarrow B', C \rightarrow C'$, then A', B', C' are collinear.

Proof: By (2.3, corollary 2), $C \leq (A, B)$. Also $t^{-1}At = A', t^{-1}Bt = B'$, $t^{-1}Ct = C'$. Thus $C' = t^{-1}Ct \leq t^{-1}(A, B)t = (t^{-1}At, t^{-1}Bt) = (A', B')$. herefore, A', B', C' are collinear. Dually, we have

(2.16) If L_1, L_2, L_3 are three concurrent lines and $t: L_1 \rightarrow L_1', L_2 \rightarrow L_2'$, $L_3 \rightarrow L_3'$ then L_1', L_2' and L_3' are concurrent.

(2.17) If A is on L and t: $A \rightarrow A'$, $L \rightarrow L'$, then A' is on L'. For $t^{-1}At = A', t^{-1}Lt = L'$ and $A \leq N(L)$. Thus $t^{-1}At \leq t^{-1}N(L)t$. From $t^{-1}Lt = L'$, $t^{-1}N(L)t = N(L')$ (c.f. proof of (2.2)). Therefore, $A' = t^{-1}At \leq t^{-1}N(L)t = N(L')$.

The theorems proved up to now show that the elements of G preserve the relations of incidence. We next define perspectivities and establish their usual properties.

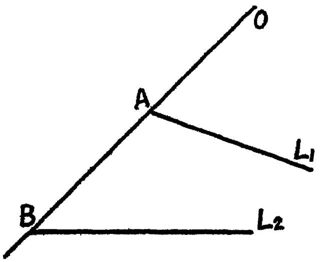
DEFINITION OF RESTRICTED PERSPECTIVITY

Let O be a point and let L_1 and L_2 be two lines neither of which passes through O. By axiom 1, $L_2 = p^{-1}L_1p$ and hence $N(L_2) = p^{-1}N(L_1)p$. Since $O \not\leq p^{-1}N(L_1)p$, we have by axiom 6, that $O \wedge N(L)p$ is not null. Let a be in $O \wedge N(L_1)p$. Put $a = kp$ where k is in $N(L_1)$. Then $a^{-1}L_1a = (kp)^{-1}L_1kp = p^{-1}(k^{-1}L_1k)p = p^{-1}L_1p = L_2$. Put $O \wedge L_1 = h_1$ and $O \wedge L_2 = h_2$. Now

$$a^{-1}h_1a = a^{-1}(O \wedge L_1)a = (a^{-1}Oa \wedge a^{-1}L_1a) = O \wedge L_2 = h_2.$$

Therefore, the cosets h_1a and ah_2 are equal. We define either of these cosets to be a *restricted* perspectivity from L_1 to L_2 with centre at O and denote it by the symbol $L_1 \overset{O}{\underset{\wedge}{\rightarrow}} L_2$ (i.e., $L_1 \overset{O}{\underset{\wedge}{\rightarrow}} L_2 = h_1a = ah_2$).

(2.18) Let t be in $L_1 \overset{O}{\underset{\wedge}{\rightarrow}} L_2$. Then:



(1). t: $O \rightarrow O$

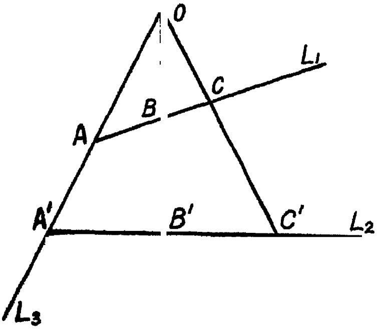
(2). t: $L_1 \rightarrow L_2$

(3). If A is on L_1 and t: $A \rightarrow B$ then B is on L_2 and A, B, O are collinear.

Proof: By definition $t = qa$ where q is in h_1 and $a: L_1 \rightarrow L_2$, and a is in O. Since q is in h_1 and $h_1 = (O \wedge L_1)$ q is in O.

Thus qa is in O. Therefore qa: $O \rightarrow O$. Again $t^{-1}L_1t = (qa)^{-1} L_1(qa) = a^{-1}(q^{-1}L_1q)a = a^{-1}L_1a = L_2$. Hence a: $L_1 \rightarrow L_2$. If A is on L_1 and $t: A \rightarrow B$ then B is on L_2 by (2.17). Also $t: O \rightarrow O$, $A \rightarrow B$, i.e.,

$t^{-1}Ot = O, t^{-1}At = B$. Put $(O, A) = N(L_3)$ and $(O, B) = N(L_4)$ (axiom 4). Then $t^{-1}N(L_3)t = t^{-1}(O, A)t = (t^{-1}Ot, t^{-1}At) = (O, B) = N(L_4)$. But t is in O and $O \leq N(L_3)$, hence t is in $N(L_3)$ and by axiom 2, t is in $NN(L_3)$. Therefore, $t^{-1}N(L_3)t = N(L_3) = N(L_4)$ or by (2.2) $L_3 = L_4$ Thus O, A, B are collinear on L_3.



(2.19) Let L_1 and L_2 be two lines and O a point not on either. Let A, B, C be three points on L_1; A', B', C' be three points on L_2 such that AA', BB', CC' pass through O. Suppose t is an element of G such that t: $O \rightarrow O$, $A \rightarrow A'$, $B \rightarrow B'$, $C \rightarrow C'$.

Then t is in $L_1 \xrightarrow[\wedge]{O} L_2$.

Proof: $t^{-1}Ot = O, t^{-1}At = A', t^{-1}Bt = B', t^{-1}Ct = C'$. By axiom 4 and (2.2) $(O, A) = (O, A') = N(L_3)$ (say). Also

$$t^{-1}N(L_3)t = t^{-1}(O, A)t = (t^{-1}Ot, t^{-1}A\,t = (O, A') = N(L_3).$$

Thus t is in $NN(L_3)$ or t is in $N(L_3)$ (axiom 2). Hence t keeps the line AA^1 fixed. Similarly t keeps BB' and CC' fixed. Thus t is in O (2 11). Also t: $L_1 \rightarrow L_2$ (2.15—2.17). Let a be in O such that a: $L_1 \rightarrow L_2$. (The definition of restricted perspectivity includes a proof of the existence of a.) Consider ta^{-1}. We have $(ta^{-1})^{-1}A(ta^{-1}) = a(t^{-1}At)a = aA'a^{-1} = A$. Thus ta^{-1} keeps A fixed. Similarly, ta^{-1} keeps B and C fixed. By (2.12) ta^{-1} is in L_1. But ta^{-1} is in O. Therefore ta^{-1} is in $O \wedge L_1 = h_1$. Hence t is in $h_1a = L_1 \xrightarrow[\wedge]{O} L_2$.

Corollary. $L_1 \xrightarrow[\wedge]{O} L_2$ consists of all transformations t of G, such that t: $O \rightarrow O$ and for each A of L_1, t: $A \rightarrow A'$ where A' is on L_2 and O, A, A' are collinear. We have used the adjective restricted in connection with the above type of perspectivity since such a perspectivity keeps the centre fixed. We now define a more general type of perspectivity.

DEFINITION OF GENERAL PERSPECTIVITY

By definition $L_1 \xrightarrow[\wedge]{O} L_2 = h_1a = ah_2$. We define a general perspectivity from L_1 to L_2 with centre at O as $L_1 \xrightarrow{O} L_2 = h_1aL_2 = ah_2L_2$. Since $h_2 = O \wedge L_2$, $h_2 \leq L_2$. Thus $ah_2L_2 = aL_2$. When there is no possibility of confusion the term general perspectivity will be contracted to perspectivity.

(2.20) If t is in $L_1 \xrightarrow[\wedge]{O} L_2$ then for any A on L_1, t: $A \rightarrow A'$ where A' is on L_2 and A, A', O are collinear.

Proof: $t = kl_2$ where k is in $L_1 \xrightarrow[\wedge]{O} L_2$ and l_2 is in L_2. Now $k: A \rightarrow A'$ where A' is on L_2 and A, A', O are collinear. Also $l_2: A' \rightarrow A'$. Thus $t: A \rightarrow A'$.

(2.21) Suppose A, B, C are three points on L_1 and A', B', C' are three points of L_2 such that AA', BB', CC' pass through O. If t: $A \rightarrow A'$, $B \rightarrow B'$, $C \rightarrow C'$ then t is in $L_1 \xrightarrow{O} L_2$.

Proof: Let k be in $L_1 \xrightarrow[\wedge]{O} L_2$. Then k: $A \rightarrow A'$, $B \rightarrow B'$, $C \rightarrow C'$. Thus $k^{-1}t$: $A' \rightarrow A'$, $B' \rightarrow B'$, $C' \rightarrow C'$. Hence k^{-1} t is in L_2 (2.12). Therefore, t is in kL_2, i.e., t is in $L_1 \xrightarrow{O} L_2$.

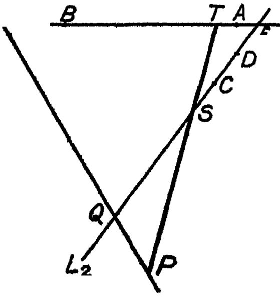
Corollary. $L_1 \xrightarrow{O} L_2$ consists of all elements of G which carry the points of L_1 into the points of L_2 in such a way that any point and its transformed point are collinear with O. This is the justification for the use of the term general perspectivity.

(2.22) Let t be an element of G such that for every point P of L_1, t: $P \rightarrow P'$ where P' is on L_2. Let L_1 and L_2 meet at A. If t: $A \rightarrow A$ then there is a point O such that t is in $L_1 \xrightarrow{O} L_2$.

Proof: Let B and C be two points of L_1 and let t: $B \rightarrow B'$, $C \rightarrow C'$, where B' and C' are on L_2. Let BB', CC' meet in O (by 2.4). Let k be any element of $L_1 \xrightarrow[\wedge]{O} L_2$. Then k: $A \rightarrow A$, $B \rightarrow B'$, $C \rightarrow C'$. From (2.7, and (2.8), $k^{-1}t$ keeps fixed A, B', C'. Hence $k^{-1}t$ is in L_2 or t is in kL_2, i.e., t is in $L_1 \xrightarrow{O} L_2$.

Definition. Let t be any element of G and L_1 be any line. Let $t^{-1}L_1t = L_2$. To each A of L_1, t: $A \rightarrow A'$ where A' is on L_2. Thus the element t induces a correspondence between the points of L_1 and the points of L_2. We define such a correspondence to be a projectivity. Using this definition (2.22) may be reworded to read: a necessary and sufficient condition for a projectivity t between L_1 and L_2 to be a perspectivity, is that t keeps the point common to L_1 and L_2 fixed.

(2.23) Let A, B, C, D be four points no three of which are collinear. If an element t of G keeps A, B, C, D fixed, t is the identity.



Proof: Let AB meet CD at E. Denote by L_1 and L_2 the lines AB and CD respectively. Since t: $A \rightarrow A$ and $B \rightarrow B$, t: $L_1 \rightarrow L_1$ (2.13). Similarly t: $L_2 \rightarrow L_2$. Thus t: $E \rightarrow E$. Hence t keeps fixed each point of L_1 and each point of L_2. Let P be a point not on L_1 or L_2. Draw two lines PST and PQR to meet

L_1 and L_2 at R, Q, S, and T as in the diagram. We have $t: S \to S$, $T \to T$, so that t keeps fixed the line PST. Similarly, t keeps fixed the line PQR. Thus t keeps the point P fixed by (2.14). Therefore, t is in $N(P)$ for all points P. Let L be any line through three points U, V, W. Then $L = N(U) \wedge N(V) \wedge N(W)$ by axiom 5. Therefore, t is in L for all L. Similarly t is in P for all points P. Therefore, by axiom 1, $t = 1$.

COROLLARY. If an element t of G keeps all points fixed t is the identity.

(2.24) Let A, B, C, D; A', B', C', D' be two sets of four points no three of either set being collinear. There exists at most one transformation t of G such that $t: A \to A'$, $B \to B'$, $C \to C'$, $D \to D'$.

Proof: If t_1 and t_2 were two such transformations $t_1t_2^{-1}: A \to A$, $B \to B$, $C \to C$, $D \to D$. Hence by (2.23), $t_1t_2^{-1} = 1$ or $t_1 = t_2$.

On strengthening axiom 6, we will show that the element t of (2.24) always exists.

(2 25) If A and B are two distinct points $A \wedge B = 1$.

Proof: Let P be any point not on AB and let $PA = L_1$, $PB = L_2$. Let k be in $A \wedge B$. Then by (2.10) k keeps L_1 and L_2 fixed and so $k: P \to P$. If Q be any fourth point such that no three of A, B, P, Q are collinear the above proof shows that $k: Q \to Q$. Hence, by (2.23), $k = 1$. Therefore $A \wedge B = 1$. The dual may be proved in the same way yielding:

(2.26) If L_1 and L_2 are two lines $L_1 \wedge L_2 = 1$.

It is now quite easy to obtain proofs of Pappus' and Desargues' Theorems, but as these are not essential to our general argument they will be omitted here. A proof of the fundamental theorem of projective geometry will be given instead.

(2.27) The points and lines of G form a projective plane.

This theorem follows from the fact that all the axioms of projective geometry (A, E, P in the notation of Veblen and Young) are theorems of our system.

(2.28) Let L be any line and P be any point. Let A and B be two points collinear with P but not on L. There is at most one element k in $L \wedge P$ such that $k: A \to B$.

Proof, Case 1: P is not on L. Let PAB meet L at C and denote the line PAB by L_1. Let k and t be two elements of $P \wedge L$ such that $k: A \to B$, $t: A \to B$. Then kt^{-1} is in $P \wedge L$ and $kt^{-1}: A \to A$. Also $kt^{-1}: P \to P$, $C \to C$ since kt^{-1} is in $P \wedge L$. Thus kt^{-1} is in L_1, and therefore is in $L \wedge L_1 = 1$. Hence $k = t$.

Case 2: Suppose P is on L. Let U be any point not on L or L_1. Let UA meet L at V. Choosing k and t as in case 1, we have kt^{-1}: $A \to A$, $V \to V$, $P \to P$. By (2.13) kt^{-1} keeps VA fixed and since kt^{-1} is in P, it keeps the line UP fixed. By (2.14) kt^{-1}: $U \to U$ and using (2.23, corollary) $kt^{-1} = 1$ or $k = t$. The dual is proved in the same way to yield the following theorem:

(2.29) Let L be any line and P be any point. Let M and N be two lines concurrent with L neither of which pass through P. Then $P \wedge L$ contains at most one element t such that t: $M \to N$.

We now replace axiom 6 by the following axiom which is at least as strong (i.e., we will show that the new axiom together with axioms 1, 2, 3, 4, 5 implies axiom 6).

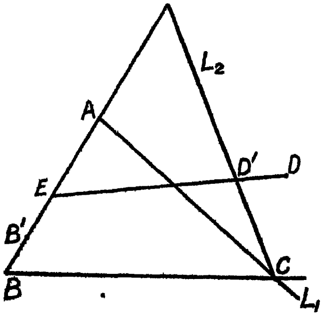
Axiom 6′. The element k of (2.28) exists.

(2.30) (Dual of axiom 6′). Let P and L be a point and line respectively. Let M and R be two lines concurrent with L neither of which passes through P. Then there exists exactly one element k of $(L \wedge P)$ such that k: $M \to R$.

Proof: It is immaterial whether P is on L or not. Draw any line through P not passing through the point of concurrence K of L, M, R and let it meet M and R at A and B respectively. By axiom 6′ there is an element k in $(P \wedge L)$ such that k: $A \to B$. By (2.3, corollary 1), $(A, K) = N(M)$ and $(B, K) = N(R)$. $k^{-1}N(M)k = k^{-1}(A, K)k = (k^{-1}Ak, k^{-1}Kk) = (B, K) = N(R)$. Let $k^{-1}Mk = S$; then $k^{-1}N(M)k = N(S) = N(R)$. By (2.2) $R = S$. Hence, $k^{-1}Mk = R$ or k: $M \to R$. That k is unique follows from (2.29).

Proof of axiom 6: Let O be any point and L_1, L_2 be two lines not through O. Let L_3 be any line, different from L_1 and L_2 which passes through their intersection. By (2.30) there exists an element t in $O \wedge L_3$ such that t: $L_1 \to L_2$. Therefore, $t^{-1}L_1t = L_2$. Let g be any element such that $g^{-1}L_1g = L_2$. Then $t^{-1}L_1t = g^{-1}L_1g$ or $gt^{-1}L_1tg^{-1} = L_1$, i.e., tg^{-1} is in $N(L_1)$ or t is in $N(L_1)g$. But t is in O. Therefore t is in $O \wedge N(L_1)g$.

(2.31) Let A, B, C, D be four points no three of which are collinear. Let D^1 be any point not on AB, BC or CA. There exists an element t of G such that t: $D \to D^1$ and t keeps A, B and C fixed.



C4a

Proof: Let DD' meet AB at E. Denote by L_1 and L_2 the lines AC and CD' respectively. By axiom 6′, there exists a unique element g in $L_1 \wedge E$ such that

$q: D \to D'$. Then $q: A \to A$, $C \to C$, $D \to D'$, $B \to B'$ where B' is on the line EB and is distinct from A and E. Take r in $(L_2 \wedge A)$ such that $r: B' \to B$ (r exists by axiom 6′). Then r: $A \to A$, $C \to C$, $B' \to B$, $D' \to D'$. Then $t = qr$ keeps A, B and C fixed, and makes $D \to D'$.

Corollary. The element t is unique, for if u be any other such element tu^{-1} keeps A, B, C, D fixed and hence by (2.23) $tu^{-1} = 1$ or $t = u$.

(2.32) Let A, B, C, D and A', B', C', D' be two sets of four points no three of which in either set are collinear. (The points A', B', C', D' are not necessarily distinct from the points of the set A, B, C, D.) There exists a unique element t of G such that t: $A \to A'$, $B \to B'$, $C \to C'$ and $D \to D'$.

Proof: Using (2.24) it is only necessary to prove that at least one such t exists. In order to avoid the many cases which arise owing to the possibility that some or all of A', B', C', D' may lie on the sides of the complete quadrangle A, B, C, D, we take another set of four points A'', B'', C'', D'' no three of which are collinear and none of which lie on the sides of either of the complete quadrangles formed by A, B, C, D and A', B', C', D' respectively. By (2.32) there exist elements a, b, c, d, of G such that

$$\begin{array}{llllllll}
a: & A \to A & b: & A \to A & C: & A \to A & d: & A \to A'' \\
 & B \to B & & B \to B & & B \to B'' & & B'' \to B'' \\
 & C \to C & & C \to C'' & & C'' \to C'' & & C'' \to C'' \\
 & D \to D'' & & D'' \to D'' & & D'' \to D'' & & D'' \to D''
\end{array}$$

The element $u = abcd$ sends $A \to A''$, $B \to B''$, $C \to C''$, $D \to D''$. Similarly there exists an element v such that

$$v: \ A'' \to A', B'' \to B', C'' \to C', D'' \to D'.$$

The required element t is then taken as uv.

In (2.27) we have shown that the points and lines of G form a projective plane. Denote this plane by Π. Let the projective collineation group of Π be G'.

We now prove the main theorem:

(2.33) The groups G and G' are isomorphic.

Proof: Every element of G induces a projective collineation of the points of Π. Thus $G \leq G'$. However, every element of G' is completely characterized by four arbitrary points no three of which are collinear and their images, which must form a set of four points no three of which are collinear. By (2.32) $G' \leq G$. Therefore $G = G'$.

PART III: THE AUTOMORPHISMS OF G

As an application of the results of part II, we will give a proof of the following result—the group of all automorphisms of the projective collineation group is isomorphic to the group of all collineations and correlations of the underlying geometry. We will give an explicit proof in the two-dimensional case only. Extension to n-dimensions can be carried out by use of the following properties of n-dimensional projective geometry. To each point of our n-dimensional space associate the group of all projective collineations each of which keeps fixed every hyperplane through that point. The following three properties of points hold:

1. For any two points P, Q; $P \wedge Q = 1$.

2. The point P_r is on the hyperplane determined by $P_1, P_2, \ldots P_{r-1}$ if and only if $P_r \leq (P_1, P_2, \ldots P_{r-1})$.

3. The group of a point P consists of homologies and elations exclusively. If furthermore it is shown that every automorphism of the projective collineation group carries a harmonic homology into a harmonic homology, the proofs of the two dimensional case will all carry over to the n-dimensional case in a direct manner. A proof of the above property of harmonic homologies will be given in a subsequent paper.

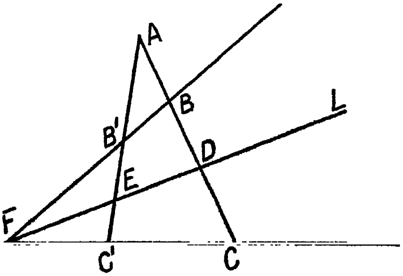
Our proofs in the two-dimensional case are synthetic, i.e., they do not depend upon the introduction of a co-ordinate field. They fail for those geometries in which a complete quadrangle has three collinear diagonal points, for in such geometries harmonic homologies do not exist. For such cases the proofs can be altered slightly by making use of projective collineations of period three. The existence of such has been established in (1.15). Analytically the exceptional cases occur when the co-ordinate field is of characteristic 2. The relevant theorems follow.

THEOREM 3.1. Let t be a harmonic homology with A as centre and L as axis. Let u be an element of G such that $uk = ku$ for each k of G such that $tk = kt$. Then u is a homology with A as centre and L as axis, and conversely every homology with A as centre and L as axis commutes with each such k.

Proof: We first show that $tk = kt$ if and only if k keeps A and L fixed. Indeed if $k: A \to A, L \to L$, $k^{-1}tk$ is a projective transformation of period 2 which keeps every point on L fixed and every line through A fixed. Hence $k^{-1}tk = t$ or $tk = kt$. On the other hand if k: $A \to \bar{A}$ and $L \to \bar{L}$ where the pair A, L is distinct from the pair

A, L, $k^{-1}tk = v$ where v is a harmonic homology with centre and axis A and L respectively. Thus $v \neq t$; therefore $tk = kv \neq kt$. If u be an element of G such that $uk = ku$ for all k, then in particular $ut = tu$, so that u: $A \to A$, $L \to L$. If u does not keep all points on L fixed suppose u: $P \to Q$ where P and Q are on L, $P \neq Q$. Let k be an element of G such that k: $A \to A$, $Q \to Q$, $P \to R$ where R is on L and $R \neq P$, $R \neq Q$. Then $k^{-1}uk$: $R \to Q$ so that $k^{-1}uk \neq u$ or $uk \neq ku$. Hence if $uk = ku$ for all k, then u keeps all points on L fixed. Similarly u keeps all lines through A fixed. Therefore, u is a homology with centre A and axis L. Suppose now that u is an arbitrary homology with A as centre and L as axis and k is an arbitrary transformation which keeps A and L fixed.

Let k $A \to A'$, $B \to B'$, $C \to C'$ where u: $B \to C$. Since A, B, D, C are collinear so are A, B', E, C'. Again $ABDC \barwedge AB'EC'$ so that BB', DE, and CC' are concurrent at F By definition u: $B' \to C'$. Now kuk^{-1}: $B \to C$ for all B or $kuk^{-1} = u$, i.e., $ku = uk$.

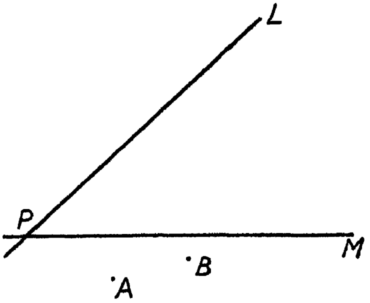


Theorem 3.2. In any automorphic mapping of G, the image of a homology is a homology and conversely, every homology is the image of some homology.

Proof: Denote the mapping by $c \to c'$ for each c of G. Let k be a homology and k' be its image. If the centre and axis of k be A and L respectively denote by t the harmonic homology with centre and axis A and L. If $t \to t'$ then t' is a harmonic homology since $t^2 = 1$ implies $t'^2 = 1$. Let the centre and axis of t' be A' and L' respectively. If $u \to u'$, then $ut = tu$ implies $u't' = t'u'$. Conversely, each element which commutes with t' is the image of some element which commutes with t. By Theorem 3.1 k' is a homology with centre A' and axis L'. By applying the inverse automorphism we see that every homology is the image of some homology.

Theorem 3.3. Let A, B be two points and L, M be two lines such that $A \neq B$ and $L \neq M$. There exist homologies k_1 and k_2 such that k_1 has centre and axis A and L, k_2 has centre and axis B and M and such that k_1k_2 is not a homology or elation.

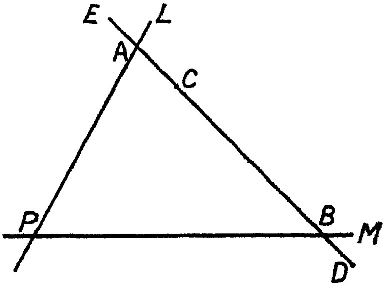
Proof, Case 1: A and B both not on L or M. In this case k_1 and k_2



may be taken arbitrarily. Let L and M meet in P. It is quite obvious that k_1k_2 has only P, and at most two points on the line AB as fixed points. Hence k_1k_2 is not a homology or elation.

Case 2: One of A or B on L or M. The same proof holds as in case 1.

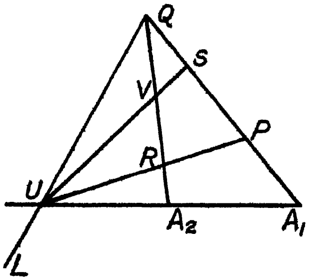
Case 3: A is on M, B on L. Let C be on AB and suppose $k_1: C \to D$. Choose k_2 such that $k_2: D \to E$, $E \neq C$. Then k_1k_2 keeps fixed A, B and P. Any other fixed point must be on the line AB. But not all points on AB are fixed since $k_1k_2: C \to E$. Therefore k_1k_2 is not a homology or elation.



Corollary. If k_1k_2 are two homologies with distinct centres and distinct axes and k_2 is not harmonic then if k_1k_2 is a homology or elation, $k_1k_2^2$ is not a homology or elation. The proof follows by an obvious extension of case 3.

Theorem 3.4. If t_1 and t_2 are two harmonic homologies with the same centre A, their images in any automorphism of G are harmonic homologies with a common centre or a common axis.

Proof: We first show that if k_1 and k_2 are two distinct homologies with the same axis (centre) and distinct centres (axes) of which k_1 is harmonic but k_2 is not harmonic, then k_1k_2 is a homology with the same axis (centre) as that of k_1 and k_2. That the product is a homology or elation with the same axis is obvious by definition. We show that it cannot be an elation. Let L be the common axis of k_1 and k_2 and let A_1 and A_2 be their respective centres. Let A_1A_2 meet L in U. If k_1k_2 is not a homology it is an elation with L as axis and U as centre. Let P be any point not on L or on the line A_1A_2. From the diagram $k_1: P \to Q$ where $H(A_1S;$



PQ) (i.e., A_1, S and P, Q are harmonic conjugates). Since k_1k_2 is an elation k_2: $Q \to R$ where R is on QA_2 and on UP (since k_1k_2 keeps UP fixed). Also $QPSA_1 \overset{U}{\underset{\wedge}{=}} QRVA_2$ so that $H(QR;\ VA_2)$. This implies that k_2 is harmonic in contradiction with our assumption. Now let $t_1 \to t_1'$, $t_2 \to t_2'$ in the given automorphism. Let the axes t_1 and t_2 be L_1 and L_2 respectively. If t_1' and t_2' have neither a common centre nor a common axis, let their centres and axes be A_1', L_1' and A_2', L_2' respectively. Let k be a homology $\neq k_2$ with centre A_2 and axis L_2. By Theorem 3.2, $k \to k'$ where k' is a homology with A_2' as centre and L_2' as axis. Now $t_1k \to t_1'k'$. By a proper choice of k, $t_1'k'$ is not a homology (Theorem 3.3). On the other hand from the remark made at the beginning of this proof t_1k is a homology. This is in contradiction with Theorem 3.2. Hence either $A_1' = A_2'$ or $L_1' = L_2'$.

Corollary 1. If k_1 and k_2 are two homologies with the same centres, their images k_1', k_2' are two homologies with a common centre or a common axis.

Proof: Let k_1 and k_2 have centres and axes A, L_1 and A, L_2 respectively, and suppose that t_1 and t_2 are the harmonic homologies which have the same centres and axes as k_1 and k_2 respectively. By Theorem 3.4, $t_1 \to t_1'$, $t_2 \to t_2'$where t_1' and t_2' have a common centre on a common axis. Since k_1' and k_2' have the same centres and axes as t_1' and t_2' respectively (Theorem 3.2), it follows that k_1' and k_2' have either a common centre or a common axis. From the principal of duality we have immediately:

Corollary 2. Any automorphism of G carries two homologies with a common axis into two homologies which have either a common centre or a common axis.

Theorem 3.5. The image of an elation in an automorphic mapping of G is an elation and conversely every elation is the image of an elation.

Proof: Let r be an elation with centre at A. Then $r = k_1k_2$ where k_1 and k_2 are homologies with a common centre A. If $k_1 \to k_1'$ and $k_2 \to k_2'$ then $r = k_1k_2 \to k_1'k_2' = r'$ under the automorphism. By (3.4, corollary 1), k_1' and k_2' are homologies with a common centre or a common axis, so that r' is a homology or an elation. If r' is a homology then the inverse mapping maps the homology r' on to the elation r, a contradiction in view of (3.2). Therefore, r' is an elation. The inverse mapping also shows that every elation is the image of an elation.

Corollary. The only sub-groups of G which consist of homologies and elations alone are points, lines, and their sub-groups (except the trivial case of a group which contains harmonic homologies only). It will be sufficient to prove that all the homologies of such a group have either a common centre or a common axis. If not, the group will contain two homologies k_1 and k_2 of which k_2 is not harmonic and such that k_1 and k_2 have distinct centres and distinct axes. By (3.3) and (3.3, corollary), either k_1k_2 or $k_1k_2^2$ is not a homology or elation. This contradicts our hypothesis.

In the subsequent work we will use our dual interpretation of points and lines. A point P is either considered as a point in a projective plane or a sub-group of the projective group G. A similar dual interpretation will be placed on lines.

THEOREM 3.6. Let P be any point. If an automorphism T makes $a \to a^T$ for each element a of G, then P^T, the group of all images of elements of P is a point or a line.

Proof: Since P consists of homologies and elations only so does P^T. By (3.5, corollary), $P^T \leq L$ where L is a point or line. If Q be any other point $Q^T \leq N$ for some point on line N. If L is a point (line), N can always be chosen to be a point (line) since P^T and Q^T are conjugate sub-groups of G, being the images of conjugate sub-groups. We first show that $N \neq L$ for if $N = L$ then $P^T \leq L$, $Q^T \leq L$ or $(P^T, Q^T) = (P, Q)^T$ is contained in L. The relation $(P, Q)^T \leq L$ is impossible since L consists of homologies and elations only whereas (P, Q) and hence $(P, Q)^T$ contain other types of collineations. Hence $N \neq L$. Again, if $P^T \neq L$ let c be an element of L not in P^T. Since c is a homology or elation it is the image of a homology or elation $\bar{c}$, where $\bar{c}$ is not in P. Since every homology or elation lies in some point $\bar{c}$ is in $Q \neq P$. If $Q^T \leq N$ then $c = \bar{c}^T$ is in N. Thus c is in $N \wedge L$. But $N \wedge L = 1$. From this contradiction we deduce that $P^T = L$ where L is a point or line.

COROLLARY 1. In the same way it can be shown that if L is a line, L^T is a point or line.

COROLLARY 2. If an automorphism T maps one point into a point it maps every point into a point and every line into a line, if T maps one point into a line it maps every point into a line and every line into a point. The proofs of both cases are the same. We prove the second. If P^T is a line for some point P, Q^T is a line for every point Q since P^T and Q^T are conjugate sub-groups of G. If L^T is a line for some line L then the mapping $P \to P^T$, $L \to L^T$ implies $P \wedge L \to P^T \wedge L^T$. But $P \wedge L$ is not the identity whereas $P^T \wedge L^T$ is the identity, a contradiction. Hence L^T is not a line and so must be a point (corollary 1).

THEOREM 3.7 (*Main Theorem*). Let G be the projective collineation group of a projective plane Π. With each automorphism of G we may associate a unique collineation or correlation of Π and, conversely, with each collineation or correlation of Π we may associate a unique automorphism of G, the mapping being an isomorphism.

Proof: Let T be an automorphic mapping of G. Consider the mapping $P \to P^T$, $L \to L^T$ as a mapping of the points and lines of Π. If P^T is a line for some point P then the mapping is a point-line, line-point mapping of Π. If P, Q, R are collinear points and P^T, Q^T, R^T are their images, the relation $P \leq (Q, R)$ implies $P^T \leq (Q, R)^T = (Q^T, R^T)$. Hence P^T, Q^T, R^T are concurrent lines. Similarly concurrent lines map into collinear points. Thus the mapping is a correlation of Π (not necessarily projective). Similarly, if P^T is a point for a point P the mapping is a collineation of Π. To prove the converse we note the following properties of G.

1. Any element of G is the product of a finite number of homologies or elations.

2. Homologies and elations, by their very construction, are projective collineations.

3. The transform of a homology or elation by any collineation or correlation of the plane Π is a transformation which has a line of fixed points and hence is a homology or elation.

Now let c be any collineation or correlation of Π and let u be any *projective* collineation. Then $u = h_1h_2\ldots h_r$ where each of the h_i is a homology or elation. Put $\bar{u} = c^{-1}uc$. Then $\bar{u} = c^{-1}(h_1h_2\ldots h_r)\, c = (c^{-1}h_1c)\ldots(c^{-1}h_rc)$. Since $c^{-1}h_ic$ is a homology or elation $\bar{u}$ is a projective collineation of G. Hence the mapping $u \to \bar{u} = c^{-1}uc$ is an automorphism of G. Finally, it is easily seen that distinct collineations or correlations of Π induce distinct automorphisms of G.

The work carried out in the above paper is part of a thesis, written by the author in April, 1942, as part of the requirements for a Ph.D. degree at the University of Toronto. The delay in publication is partly due to the war. The author wishes to express his thanks here to Professor G. de B. Robinson for very valuable aid during the carrying-out of this work.

ON A BOUNDARY VALUE PROBLEM FOR A CLAMPED PLATE

By A. WEINSTEIN and J. A. JENKINS

Presented by S. BEATTY, F.R.S.C.

1. INTRODUCTION

THE present paper provides the theoretical basis for the variational method employed in a previous note by A. Weinstein and D. H. Rock (1944). Up to the present this method has been applied only to the computation of *eigenvalues* (Weinstein, 1937). Here we deal with functions approximating to the solution of a boundary value problem for the clamped plate. We will follow the general outline of the paper "On the Unified Theory of Eigenvalues of Plates and Membranes" by N. Aronszajn and A. Weinstein (1942), which we will refer to as AW.

2. THE VARIATIONAL PROBLEMS

Let S be the domain of our plate in the (x, y)-plane. We assume that it is bounded, of finite connectivity and with a boundary consisting of a finite number of arcs of bounded mean curvature, two arcs having at most their end points in common, the angles at the common vertices being > 0 and $\leq 2\pi$.

The following notations will be employed throughout the paper:

(1) $J(w) \equiv \iint [(\Delta w)^2 - 2qw]\, dx\, dy,$

where q denotes a fixed function;

(2) $(v, w) \equiv \iint_S v\, w\, dx\, dy.$

We formulate now the variational boundary value problem of the bending of a clamped plate as follows:

Problem VP: To find a function minimizing the expression $J(w)$ among all functions w continuous in $S + C$ possessing continuous first and second derivatives in S, the Laplacian Δw being of integrable square in S. Moreover, w must satisfy the boundary conditions

(3) $w = 0$ on C;

(4) $\dfrac{dw}{dn} = 0$ on C (at points of the boundary admitting a normal).

As in the previous paper, we consider a new variational problem VP_0 obtained from VP by subjecting the minimizing function to only the first boundary condition

(3) $w = 0$ on C.

It is well known that the corresponding differential problem can be solved in terms of the problem of the equilibrium of a membrane. The solution of VP_0 will provide the first step in the approximation to the solution of VP. For this reason, VP_0 will be called the base problem.

The base problem will be linked to VP, or, more exactly, to an equivalent problem, by a sequence of intermediate variational problems introduced in the following manner:

Let

(5) $p_1(x, y), p_2(x, y), \ldots$ denote a sequence of linearly independent harmonic functions of integrable square and regular in S and complete in the set $\mathfrak{p}$ of harmonic functions of integrable square in S.

We then define the intermediate variational problems VP_m ($m = 1, 2, \ldots$) as follows:

Problem VP_m: To find a function minimizing the expression $J(w)$ among all functions w continuous in $S + C$ possessing continuous first and second derivatives in S, the Laplacian Δw being of integrable square in S.

Moreover, w must satisfy the boundary conditions:

(3) $w = 0$ on C;

(6) $(p_k, \Delta w) = 0 \; k = 1, 2 \ldots, m$.

Finally, we consider a variational problem VP_∞ defined in the same manner as VP_m, except that the condition

(6) $(p_k, \Delta w) = 0, k = 1, 2 \ldots m$,

is to be replaced by the condition

(7) $(p_\iota, \Delta w) = 0$, for all $p_\iota(x, y)$.

We will show that the solutions of VP_m can be expressed in terms of the solution of VP_0 and that they converge to the solution of VP_∞. Further, it can be shown, as in AW, that the solutions of VP and VP_∞ are identical.

3. EXISTENCE AND UNIQUENESS FOR THE PROBLEMS VP_m AND VP_∞

The Green's potential Gf of a function f of integrable square is defined by

(8) $Gf = G(x, y; f) = -\iint g(x, y, \xi, \eta) f(\xi, \eta) d\xi d\eta$,

where $g(x, y, \xi, \eta)$ is the Green's function for the domain S. The properties of this operator are outlined in AW, section 2.

For functions satisfying

(3) $w = 0$ on C

the equation $\Delta w = f$ is equivalent to the equation $w = Gf$. Hence for these functions we can write

(9) $J(w) = (f, f) - 2(Gf, q)$,

where we have set $\Delta w = f$.

Further, the problem VP_m amounts to minimizing

(9) $(f, f) - 2(Gf, q)$

under the conditions

(10) $(f, p_n) = 0, n = 1, \ldots, m.$*

Since

(11) $(f - Gq, f - Gq) \geq 0$

we have

(12) $(f, f) - 2(f, Gq) \geq -(Gq, Gq)$.

Since G is a symmetric operator the left side is equal to the expression (9), which is thus bounded below. Let μ_m denote its greatest lower bound and let $\{f^{(\sigma)}\}$ denote a sequence of functions satisfying the boundary conditions of VP_m and such that the corresponding values of the functional J converge to μ_m.

This implies

(13) $(f^{(\sigma)}, f^{(\sigma)}) - 2(f^{(\sigma)}, Gq) = \mu_m + \epsilon^{(\sigma)}$,

where $\epsilon^{(\sigma)} \to 0$ as $\sigma \to \infty$.

Hence

(14) $(f^{(\sigma)} - Gq, f^{(\sigma)} - Gq) = \mu_m + (Gq, Gq) + \epsilon^{(\sigma)}$,

and so the left-hand side is bounded above for all σ.

Moreover, by the triangle inequality,

(15) $(f^{(\sigma)}, f^{(\sigma)})^{\frac{1}{2}} \leq (f^{(\sigma)} - Gq, f^{(\sigma)} - Gq)^{\frac{1}{2}} + (Gq, Gq)^{\frac{1}{2}}$,

so that $(f^{(\sigma)}, f^{(\sigma)})$ is uniformly bounded for all σ. Thus the sequence $\{f^{(\sigma)}\}$ contains a weakly convergent subsequence which we denote by $\{f^{(\sigma)}\}$ again, which has a weak limit which we will call provisionally $\bar{f}$.

---

*This new problem, which we denote by VP_m again, might at first sight appear more general than the original one. However, the function $w = Gf$ corresponding to the solution f of the new problem will be found to have two continuous derivatives, and thus the problems actually are equivalent.

Again we have

(16) $(f^{(\sigma)}, f^{(\sigma)}) = 2(f^{(\sigma)}, Gq) + \mu_m + \epsilon^{(\sigma)}$,

and taking the limit as $\sigma \to \infty$

(17) $$(\bar{f}, \bar{f}) \leq \lim_{\sigma\to\infty} (f^{(\sigma)}, f^{(\sigma)}) = \lim_{\sigma\to\infty} \{2(f^{(\sigma)}, Gq) + \mu_m + \epsilon^{(\sigma)}\}$$
$$= 2(\bar{f}, Gq) + \mu_m.$$

We next observe that $\bar{f}$ is an admissible function in VP$_m$ since

(18) $$(\bar{f}, p_k) = \lim_{\sigma\to\infty} (f^{(\sigma)}, p_k) = 0,\ k = 1, \ldots, m.$$

Thus also

(19) $(\bar{f}, \bar{f}) - 2(\bar{f}, Gq) \geq \mu_m$,

so that finally

(20) $(\bar{f}, \bar{f}) - 2(\bar{f}, Gq) = \mu_m$.

This proves the existence of a solution of VP$_m$, which will be denoted by f_m.

The existence of a solution of VP$_\infty$ is proved in precisely the same manner.

We denote the solution by f_∞ and the corresponding minimum by μ_∞.

A necessary condition for a function h to be a solution of VP$_m$ is obtained by considering an arbitrary function ϕ orthogonal to the space $\mathfrak{p}_m$ spanned by $p_1, \ldots\ldots, p_m$. For any such ϕ

(21) $$(h - 2Gq, h) \leq (h + \epsilon\phi - 2Gq, h + \epsilon\phi)$$
$$\leq (h - 2Gq, h) + \epsilon(2h - 2Gq, \phi) + \epsilon^2(\phi, \phi).$$

For h to provide an actual minimum the coefficient of ϵ must vanish, i.e.

(22) $(h - Gq, \phi) = 0$,

for all functions ϕ orthogonal to the space $\mathfrak{p}_m$. These functions ϕ constitute a linear manifold Φ completely orthogonal to the linear manifold $\mathfrak{p}_m$. Thus by the general projection theorem

$h - Gq$ must lie in $\mathfrak{p}_m$.

Therefore almost everywhere in S

(23) $$h = Gq + \sum_{j=1}^{m} a_m^{(j)} p_j.$$

By modifying the solution in a set of measure zero we can have this equation holding everywhere in S.

Further, for h to be a solution of VP$_m$, it must satisfy the boundary conditions

(24) $(h, p_k) = 0, \; k = 1, \ldots, m,$

so that

(25) $\sum_{j=1}^{m} a_m^{(j)} (p_j, p_k) = -(Gq, p_k), \; k = 1, \ldots, m.$

These equations determine the $a_m^{(j)}$ uniquely since the determinant of their coefficients is non-zero, being Gram's determinant for the independent functions $p_1, \ldots, p_m$.

Since f_m is a solution it is by (23) and (25) the unique solution and is expressible in the form

(26) $f_m = Gq + \sum_{j=1}^{m} a_m^{(j)} p_j.$

Similarly any solution of VP_∞ must have the form

(27) $h = Gq + p,$

where p is a harmonic function in $\mathfrak{p}$.

As above we must have also

(28) $(h, p_k) = 0$ for all p_k,

so that

(29) $(p, p_k) = -(Gq, p_k)$ for all k.

By the completeness of the sequence $\{p_k\}$ in the set $\mathfrak{p}$, p is determined uniquely.

Hence, as before, since f_∞ is a solution, it is by (27) and (29) unique and is expressible in the form

(30) $f_\infty = Gq + p.$

4. CONVERGENCE

Next we shall prove that the functions f_m converge strongly to f_∞ in S.

We observe that we have for the problems $VP_m (m = 0, 1, 2, \ldots, \infty)$ minima

$$\mu_0 \leqq \mu_1 \leqq \mu_2 \leqq \ldots \leqq \mu_\infty,$$

and we know that μ_∞ is finite.

Since

(31) $(f_m - Gq, f_m - Gq) = (f_m, f_m) - 2(f_m, Gq) + (Gq, Gq)$
$\quad = \mu_m + (Gq, Gq) \leqq \mu_\infty + (Gq, Gq),$

the first term is bounded above for all m.

Moreover,

(32) $(f_m, f_m)^{\frac{1}{2}} \leqq (f_m - Gq, f_m - Gq)^{\frac{1}{2}} + (Gq, Gq)^{\frac{1}{2}}$, so that (f_m, f_m) is uniformly bounded for all m.

Thus the sequence $\{f_m\}$ contains a weakly convergent sub-sequence $\{f_{m_k}\}$ with weak limit f say.

Since

(33) $(f_m, p_k) = 0$ for $k \leq m$,

we have

(34) $(f, p_k) = 0$ for all k.

Moreover,

(35) $(f_{m_k}, f_{m_k}) - 2(f_{m_k}, Gq) = \mu_{m_k} \leq \mu_\infty$,

so that

(36) $(f, f) - 2(f, Gq) \leq \underline{\lim}_{k \to \infty} (f_{m_k}, f_{m_k}) - 2(f, Gq) \leq \mu_\infty$.

Since f is an admissible function in VP$_\infty$ this implies

(37) $(f, f) - 2(f, Gq) = \mu_\infty$,

and since the solution of VP$_\infty$ is unique $f = f_\infty$.

Indeed, the entire sequence $\{f_m\}$ must converge weakly to f_∞ for, if it did not, there would exist a sub-sequence which would converge weakly to $f^* \neq f$. However, by the same argument as above, we would have $f^* = f_\infty$ and so a contradiction.

The above proof further establishes that

(38) $\lim_{h \to \infty} \mu_{m_h} = \mu_\infty$.

Since the sequence of the μ's is monotone increasing this implies

(39) $\lim_{m \to \infty} \mu_m = \mu_\infty$.

In the relation

$$(f_m, f_m) = 2(f_m, Gq) + \mu_m$$

the right-hand side possesses a limit as $m \to \infty$. Hence so does the left-hand side, and we have

(40) $\lim_{m \to \infty} (f_m, f_m) = 2(f_\infty, Gq) + \mu_\infty = (f_\infty, f_\infty)$.

Finally,

(41) $(f_m - f_\infty, f_m - f_\infty) = (f_m, f_m) + (f_\infty, f_\infty) - 2(f_m, f_\infty)$.

As $m \to \infty$ the right-hand side has limit zero, thus

(42) $\lim_{m \to \infty} (f_m - f_\infty, f_m - f_\infty) = 0$,

i.e., f_m converges strongly to f_∞, as stated.

We note further

(26) $f_m = Gq + \sum_{j=1}^{m} a_m^{(j)} p_j = Gq + h_m$ say;

(30) $f_\infty = Gq + p$,

so that h_m converges strongly to p.

It is well known that this implies $h_m(x, y)$ converges to $p(x, y)$ in S and uniformly so in every closed sub-domain of S.

To f_m and f_∞ correspond solutions w_m and w_∞ of the problems VP$_m$ and VP$_\infty$ in their original form. They are expressible as

(43) $w_m = Gf_m = GGq + Gh_m;$

(44) $w_\infty = Gf_\infty = GGq + Gp.$

For these functions to be solutions of VP$_m$ and VP$_\infty$, it is necessary that w_m and w_∞ and hence GGq possess continuous first and second derivatives. For this it is sufficient that q be bounded and of integrable square.

The first set are the equations employed in numerical computation of the approximations.

We now show that w_m converges uniformly to w_∞ in $S + C$. Indeed,

(45) $w_\infty - w_m = G(p - h_m) = -\iint_S g(x, y, \xi, \eta)\,[p(\xi, \eta) - h_m(\xi, \eta)]d\xi d\eta,$

so that

(46) $|w_\infty - w_m|^2 \leq (\iint_S g^2(x, y, \xi, \eta)d\xi d\eta)(\iint_S [p(\xi, \eta) - h_m(\xi, \eta)]^2 d\xi d\eta).$

Here on the right-hand side the first factor is uniformly bounded for (x, y) in $S + C$, and the second tends to zero as $m \to \infty$. This proves the assertion that $|w_\infty - w_m| \to 0$ uniformly.

The functions Gh_m and Gp possess all derivatives in the interior of S. We shall show that the derivatives of Gh_m converge uniformly to those of Gp in every closed sub-domain $\bar{S}$ of S. w_m and w_∞ possess as many derivatives as GGq, and all existing derivatives of w_m converge to the corresponding ones of w_∞ since

$$w_\infty - w_m = Gp - Gh_m.$$

To prove this result for the first derivative we consider a point $P(x, y)$ of $\bar{S}$ and write

(47) $w_\infty - w_m = -[\iint_{\bar{S}} + \iint_{S-\bar{S}} g(x, y, \xi, \eta)\{p(\xi, \eta) - h_m(\xi, \eta)\}d\xi d\eta].$

We can differentiate with respect to x under the integral sign each term on the right-hand side, the first because $p - h_m$ is regular in $\bar{S}$, the second because P is exterior to $S - \bar{S}$.

Thus

$$(48)\quad \frac{\partial(w_\infty - w_m)}{\partial x} = -\left[\iint_{\bar{S}} \frac{\partial g(x, y, \xi, \eta)}{\partial x} \{ p - h_m \}\, d\xi d\eta \right.$$
$$\left. + \iint_{S-\bar{S}} \frac{\partial g(x, y, \xi, \eta)}{\partial x} \{ p - h_m \}\, d\xi d\eta \right],$$

so that

$$(49)\quad \left|\frac{\partial(w_\infty - w_m)}{\partial x}\right| \leqq \epsilon_m \iint_{\bar{S}} \left|\frac{\partial g}{\partial x}\right| d\xi d\eta$$
$$+ \sqrt{\iint_{S-\bar{S}} \left(\frac{\partial g}{\partial x}\right)^2 d\xi d\eta}\ \sqrt{\iint_{S-\bar{S}} (p - h_m)^2 d\xi d\eta}\ .$$

Here ϵ_m denotes the maximum of $| p(x, y) - h_m(x, y)|$ in $\bar{S}$ and so tends to zero as m tends to infinity. Thus the right-hand side tends to zero uniformly for (x, y) in $\bar{S}$ as m tends to infinity.

Evidently the analogous result is true for differentiation with respect to y.

In the case of the higher derivatives we can use Green's theorem relative to $\bar{S}$ to express $\frac{\partial(w_\infty - w_m)}{\partial x}$, for example, as the sum of an indefinitely differentiable function and a term like the first integral in $w_\infty - w_m$ above. The only difference is that

$$\{ p(\xi, \eta) - h_m(\xi, \eta) \} \text{ is replaced by } \frac{\partial}{\partial \xi} \{ p(\xi, \eta) - h_m(\xi, \eta) \}\ .$$

This does not affect the proof since the derivatives of $p - h_m$ converge in the same manner as the function itself.

This completes the proof of all the desired results for the problem VP_∞.

5. EQUIVALENCE OF THE PROBLEMS VP AND VP$_\infty$

The equivalence of the problems VP and VP_∞ is proved precisely as in AW, section 5. (Indeed, all results proved here are equally true for the wider class of domains employed there.) Thus the existence and convergence properties proved for the solution of VP_∞ hold for the solution of VP. It should be observed that in order to obtain an independent existence proof for VP we must make use of the second proof in AW, section 5, no. 2.

6. EQUIVALENCE OF THE VARIATIONAL PROBLEM WITH THE DIFFERENTIAL PROBLEM

As in the first part of AW, section 6, we obtain at once:

The solution w of the variational problem VP is the unique solution of the differential equation

$$\Delta\Delta w = q \tag{50}$$

satisfying the boundary conditions

$$w = 0 \text{ on } C; \tag{3}$$

$$\frac{dw}{dn} = 0 \text{ on } C \text{ (at points of the boundary admitting a normal).} \tag{4}$$

REFERENCES

ARONSZAJN, N. and WEINSTEIN, A. On the Unified Theory of Eigenvalues of Plates and Membranes. Amer. Jour. Math, **64:** 623-43. 1942.

WEINSTEIN, A. Etude des spectres des équations aux dérivées partielles de la théorie des plaques élastiques. Mémorial des sciences mathématiques, no. 88. 1937.

———— and ROCK, D. H. On the Bending of a Clamped Plate. Quart. App. Math., **2:** 262-6. 1944

TRANSACTIONS OF

THE ROYAL SOCIETY
OF CANADA

SECTION III

CHEMICAL, MATHEMATICAL,
AND
PHYSICAL SCIENCES

THIRD SERIES -VOLUME XLI—SECTION III

MAY, 1947

OTTAWA
THE ROYAL SOCIETY OF CANADA
1947

TABLE OF CONTENTS

The Expanding Role of Physics in Biology and in Medicine. Presidential Address, by E. L. Harrington, F.R.S.C. . . 1

On a Conjecture by Nakayama

By Richard Brauer, F.R.S.C. . . 11

By G. de B. Robinson, F.R.S.C. 20

Oscillating Satellites with the Force Varying Inversely as the nth Power of the Distance. By Daniel Buchanan, F.R.S.C. . 27

Illumination for Concealment of Ships at Night. By E. Godfrey Burr 45

Transverse Waves in Plane Soap Films. By John Satterly, F.R.S.C. 55

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLI : SERIES III : MAY, 1947
SECTION THREE

PRESIDENTIAL ADDRESS

The Expanding Role of Physics in Biology and in Medicine

By E. L. HARRINGTON, F.R.S.C.

TO present vividly the revolutionary effect on biology and medicine of the application of the principles and instruments of physics in these fields one would need to recall to your minds the state of medicine during the centuries when the scientific method, typical of physics, had little place. In those days medicine was considered a secret art and so-called physicians or medicine men employed strange incantations and magic brews of most ridiculous and even revolting substances. They guarded well their mysteries and magical formulae.

A start towards scientific medicine was made when physics gave to the physicians the microscope and thereby made possible direct observation of many life processes in body cells. The science of physiology is said to have begun with Harvey, who was among the first to apply systematic measurements to biological phenomena and to point the way to advances in our knowledge of circulation through the use of manometers, the stethoscope, and the microscope. But the advance was slow, partly because of the opposition of the church, the state, and the medical men themselves who did not want to have their magic subjected to the acid tests of science. Indeed the study of anatomy and the applications of physical apparatus to the human body, or even to experimental animals, were considered a crime against God and man. During the centuries through which such attitudes were dominant medicine advanced but slowly and at times lost ground.

But science gradually acquired freedom during the last century and in the lifetime of the senior men of science and of medicine now living the world has witnessed a tremendous growth in all sciences, particularly in physics, and perhaps as much progress in what may be termed scientific medicine as in all previous time. Medicine has become a science rather than a mere art, and there is a growing appreciation of the flood of new principles, new methods, and new instruments which have been made available as a direct contribution of the sciences.

Physics is generally recognized as the basic science, and its principles form the foundations not only for chemistry, geology, meteorology, and all branches of engineering, but just as truly for biology and medicine. Yet until recently physics has been discounted by practicing physicians and thoroughly disliked by medical students. This anomalous situation has been due neither to the poor quality of instruction nor to any incapacity on the part of physics to render valuable service to medicine. It has been due mainly to the failure of the medical men to receive in their premedical training a sufficient grasp of physics to enable them to use it as a tool or even to realize its potentialities in their field. But the responsibility for this failure must be placed on the physicists who gave instruction in physics to premedical students, for most of them knew so little of medicine and the medical sciences that they were incapable of drawing on these fields for illustrative material. Practically all texts in physics turn to the fields of engineering for their applications of principles, and there is little or no carry-over into anatomy, physiology, or even into medical practice. The premedical student generally finishes his course in physics with a sigh of relief rather than with the sense of having acquired something which should function increasingly in his life-work. One of the texts most widely used at the present time was found to contain eleven times as many illustrations drawn from the field of engineering as from the medical sciences or from medical practice, and fifteen times as many problems. Whether this reflects ignorance or merely indifference one might hesitate to say, but it is most certainly not due to any lack of material in the latter fields.

Many of you may know that Glasser, a physicist of the Cleveland Clinic, with the help of about 250 contributors, has issued a treatise of Medical Physics containing about 1,750 large, closely printed pages. Its index of authors to whom references are made in the treatise contains over five thousand names. The fact that such a book could be written, that such a mass of literature on the applications of physics to biology and to medicine does exist, in itself is strong support for the theme of this address—that there has been a rapid extension of physics into these fields. The fact that such material is now at hand places the responsibility of giving more suitable courses in physics to the premedical students squarely on the shoulders of the professors of physics. It is now possible to give to such students physics courses quite as well suited to their needs as the courses traditionally given have been to the needs of the engineering student.

In support of this basic contention a few illustrations will be given in each of several main branches of physics by way of evidence of the fundamental importance of this science to medicine, and of the new advances in medicine made possible by the phenomenal activity in physics during the last half-century.

One may find interest even in the old field of mechanics, which has been the handmaiden of medicine and the basis of physical therapy throughout history. Yet only too recently has the analytical study of forces been carried over to the study of bone stresses, tendon and muscle tensions, and of the elastic constants of these materials. These matters are of basic importance in the rebuilding of the thousands of bodies broken in the world war. A physician who understands the elementary principles of mechanics can see stresses in the various joints, and in other parts of the skeleton, and in the muscles, which are of far greater magnitude than could be conceived intuitively. One who realizes that the normal stresses in an ankle joint may rise to several times the entire weight of the body of which it forms a part deals more intelligently with that member than one whose knowledge is qualitative rather than quantitative. His treatment of fractures is made far simpler, and the results are better and more quickly obtained than those which would be secured by one following the older practices. Many of the methods employed in the now well-established physical therapy have grown out of the revival of interest in such commonplace matters as stresses, torques, and energy considerations, as applied to the human body. Not only in Canada and the United States but in Europe as well the mechanical aspects of physical therapy are proving so effective that medical men have come to use such techniques to a degree unknown a few years ago. Yet mechanics has been one of the most unpopular branches of physics among premedical students, who are seldom led to associate mechanics with their own bodies.

Aside from the cancer problem the matter of blood circulation is the top-ranking issue before the medical fraternity today and, it might be added, a matter of greatest concern to insurance companies. Indeed the latter have joined in supporting many research projects designed to discover if possible the principal causes of failure in circulation and therefore of deaths from this cause. The work of the heart and its likelihood of failure depend primarily on the ordinary principles of hydrodynamics, on factors which pertain to the flow of liquids in general. The physiologist and the physicist find the circulatory system to be essentially a dual system of tubes with two pumps, and that the same relations among such factors as pressures, fluid velocities, tube diameters, viscosity, etc., hold as would require consideration in dealing with the problems of flow in, say, an oil refinery. In its complete circuit the blood must pass through two pumps for exactly the same reason that pumps are distributed along the course of an oil line; this makes for lower pressures along the vessels than would be possible otherwise, besides offering better controls.

The old-school physician was, or is, prone to speak qualitatively of the "strain on the heart," but the present physicians and physiologists calculate the energy requirements in terms of the pressure against which the heart must pump, the resistance offered by the orifices, and the volume and the velocity of the blood pumped into the arteries per beat; all involving an expression of several algebraic terms, and reminiscent of problems in physics. Manometers have been connected to many different parts of the circulatory system, and flow meters inserted in physiological research. The specialist not only measures the pressure but he may, by the intricate tools of physics, secure electrocardiograms by which he is guided in his diagnosis and his treatment of the patient. Lately highly complicated radio tube circuits have been developed which make possible ultra-high amplifications of the variations in electrical potentials in various parts of the body, and which have thereby thrown much light on the many vital problems related to the circulation of the blood as well as on other problems. The specialist in this field needs to know not only the simpler mechanics pertaining to pumps and to tubular systems but he must call upon the instruments having to do with hydrostatics and hydrodynamics, and he must employ also the most modern of electronic devices. Yet it is a rare text in physics that even mentions the circulation of blood in the chapter on hydrodynamics; and the professor of physics who makes the needed transfer of the principle is almost equally as rare. Yet, on this topic alone, the compendium on Medical Physics referred to above has enough material to fill perhaps fifty pages of an ordinary text.

Even though it may be a digression, a concrete example of what physics can contribute to the problem of circulation may be of interest. Every physician knows how very difficult it is to prevent bed ulcers in the case of any patient kept immovable with respect to his bed for long periods, whether because of a complicated fracture or some other cause. Any person who as a patient has experienced such an affliction will remember vividly the excruciating pain they caused, and may carry scars as mute evidence. Yet the problem has been attacked by simple mechanical means with apparently complete and definite success. In each of the several schemes which have been tried the object is the same, and the methods are similar. Each provides a motorized mechanical means of shifting periodically the resultant gravitational force on the body of the patient. This prevents blanching pressures from remaining in any area for periods long enough to cause this disintegration of the tissues under pressure. Here again clinical results from such simple means are not only startling but measure up to the predictions of physics just as completely as the output of any ordinary machine is in line with established physical principles.

The Western world has adopted with astonishing quickness the marvellous aids to hearing made possible through modern electronics. These may prolong the capacity to hear for even decades. For centuries man has employed spectacles as an aid to seeing. In each case the instrument is a gift from physics. Need one then be surprised that a mechanical aid to circulation should be found helpful in this case, where failures or deficiencies account for a large percentage of the deaths of senior persons.

Only recently have either the physicists or the physiologists attached much significance to the matter of surface tension. The work by Langmuir and his associates in America, and by Adams and a number of others in England, as well as by physicists in other countries, has led to the expansion of surface tension and absorption, a closely allied field, into a new branch of the science now called surface physics. Discoveries in this field form the bases of many entirely new industrial processes, and make possible a more complete understanding of many problems such as those of digestion and nutrition, cell growth and division, surface energy, gland secretion, enzyme and catalytic action, collodial states, the action of drugs on specific tissues, cells, or organs, the effectiveness of antidotes, the sizes and shapes of complex molecules, protein films, fatty acids and sterols, hormone action and immunology, coagulation, and the like. The present-day practices based on the new knowledge in these fields are a direct outgrowth of the contributions from surface physics. The origin of electrical potentials in the body, nerve and muscular action are viewed from a new approach. The action of medicines has always been a subject of mystery. How a minute amount of a particular substance could affect a body so large in comparison could hardly be explained except on the basis of selective adsorption. Materia Medica actually takes on new importance and requires a new approach. The potency of drugs becomes a matter of physics, rather than of magic. The premedical student who in his physics class never learns of the principles of surface physics will hardly be able later to understand many of life's fundamental processes, or even to follow up reports on new developments based on these principles.

There have been important developments in the therapeutic uses of radiant heat, of ultra-violet radiations, in the equipment for diathermy, etc., but they are well known and cannot be given adequate treatment in a brief address. Only the "sterilamp," a recent development, will be mentioned.

Surgeons and attendants have for decades gone to great pains to sterilize their instruments, their persons, and their clothing, and of course all materials associated with the operating room. But it remained for a physicist, studying vapour discharge lamps, to develop a lamp that is capable of sterilizing the air itself in an operating room. This development required the joint efforts of the physicist, the biologist, and the electrical engineer. It involved the determination of the most bactericidal wave-lengths of the spectrum, the production of a glass that would transmit the desired radiation, and the making of a lamp that would operate properly and efficiently under ordinary practical conditions. The final result of this teamwork of men in different fields is a great reduction in the occurrence of what has been termed "post-operational infection."

In the entire history of science there has never been any development so rapid or so important as the recent development of electronic tubes and tube circuits. The incredibly high amplification of minute changes in electrical potentials makes them just as useful in studying changes in potentials in different parts of the body as in the detection and reproduction of very distant broadcasts. The development of electrocardiographs of unprecedented sensitivity has been mentioned. High amplifications have made possible an almost new branch of medicine, that of electro-encephalography, more conveniently described as the study of brain waves. The electro-encephalograph offers an amazingly successful approach to the study of the action of the brain and its pathological conditions. The electrical activity of the brain was not even known prior to the work of Caton of England in 1874. The lack of sufficiently sensitive instruments blocked any progress until thirty years later when Einthoven invented the string galvanometer, and thereby made it possible to study brain waves in the lower range of frequencies. But modern electro-encephalography really dates from 1934 when the use of the electronic amplifiers established among other facts that the brain of man has a definite beat coming directly from neurones and changing with age and sensory stimulation, with disease, and with the physico-chemical state of the body; and that electro-encephalograms show beats due to combinations of sinusoidal fluctuations in voltage having frequencies of from 1 to 60 vps, with amplitudes up to say 50 microvolts, with a distinctive rhythm of about 10 vps. Recording instruments have been developed which are capable of giving "cortex frequency spectra." Certain wave forms are thirty times as frequently given by epileptics as by normal persons and their presence in an electro-encephalogram has a definite diagnostic value.

Electro-encephalograms are of clinical value with respect to epilepsy, brain tumours, trauma and haemorrhage, cerebral thrombosis, brain abscess, meningitis, encephalitis, etc.—an important, though greatly abbreviated, list. In each case they offer not only our best basis for diagnosis, but through the use of multiple channels our best means of localization. Here is a very important medical field which had to await the development by the physicist of amplifiers, oscilloscopes, and associated instruments capable of recording the intricate oscillations in potential characteristic of brain waves.

X-rays and radium, both discovered just prior to the beginning of the present century, are only now recovering from the disrepute into which they fell during the first two decades of their use in medicine. The incurable "burns" were but the direct result of the handling of powerful sources without either the knowledge or the instruments required to make intelligent and successful use of these two most potentially valuable therapeutic agents available for the treatment of cancer. Medical practitioners employed them with a readiness which far outstripped their zeal to learn the physics basic to their production and to discover their properties, as well as their limitations and dangers. The result was a baffling mixture of striking successes and tragic failures. There are still prominent practitioners using these powerful agents who do not even pretend to measure the doses given. They cannot therefore keep dependable records nor correlate results obtained with the doses given. Actually there are few fields in which quantitative controls are so important. Insufficient radiation of a malignant tumour may serve only to stimulate it and increase its resistance, and excessive radiation may destroy even healthy tissues and leave a condition less tolerable than the initial one. Physicists have provided integration meters and protective screens but few clinics are provided with a physicist whose duty it is to calculate on an integration basis the total radiation absorbed in each part of the tumour and also in the other parts of the body of the patient. The problem is particularly complicated when a multiple radioactive source is employed, but its solution spells the difference between success and the caprices of chance. Indeed success in this field has awaited the complete co-operation between the physician and the physicist which only now is being obtained. An outstanding example of such co-operation is to be found in Saskatchewan where a physicist, Professor H. E. Johns, spends half of his time in the cancer clinics, giving the radiologists appropriate lectures on the basic theory of the radiations employed, and developing methods of making essential calculations and also co-operating in making them.

This discussion should not be closed without at least a reference to the most recent contributions of physics to medicine, the artificially produced radioactive substances, and the electron microscope, developments in which Canadians have played an important part.

Of the roughly 450 new atoms which co-operating physicists and chemists have been able to produce through the use of the cyclotron and other physical means of obtaining high-speed particles there are a few, perhaps twenty, which have half-life values and types of radiation which make them suitable to act as tracers in the human body as well as in plants. In brief, it is now possible to trace the path normally taken by an inactive material, say phosphorous, in the body until it becomes appropriated in some particular tissue or organ. One has merely to mix with the inactive material a minute quantity of radioactive isotope, which in the case given would be a radioactive phosphorous, P^{32}. This thereafter accompanies the ordinary phosphorous, without change in proportions, wherever the latter may move in the body. Since P^{32} has a half-life period of 14.2 days it gives the experimentalist sufficient time for observation, and yet its life is short enough that there is no danger of overdose by prolonged exposure. Its presence in any organ can be detected by its beta radiation. In the case of wheat plants, or in the case of the skeletons of experimental rats, autoradiographs giving not only the location but also the relative concentration of the radioactive tracer, may be obtained by the direct application of the specimen under study to the photographic film. Not only P^{32} but radioactive sodium Na^{24} and some others promise to be extremely useful in a direct attack on cancer in addition to their value as tracers. Either Na^{24} or P^{32} may be introduced directly into the blood stream and the whole body thereby conveniently given the benefits of the radiation. P^{32} is found to concentrate in, and therefore to irradiate, leukemic tissues. Radioactive cobalt Co^{60} may become a more suitable source than radium itself in the treatment of cancer. Radioiodine and radiostrontium have both been found useful therapeutically as well as in tracer studies. President Conant and others of Harvard have already obtained highly significant results using radioactive carbon, C^{14}, a substance which promises to be extremely useful in the future because of its long life and the biological importance of carbon. His experiments threw an entirely new light on the question of carbohydrate metabolism. They yielded definite evidence that the whole structural make-up of the fundamental units of the body is in dynamic equilibrium.[1]

In brief the production of artificial radioactive substances has opened up a new and important field that could never have been touched by the ordinary methods of either physics or chemistry. Yet even with an abundant supply of artificial radioactive substances at hand tracer experiments as we know them would be impossible without the use of the versatile Geiger counter, another gift from physics, an instrument immeasurably more sensitive as a detector than any chemical method known.

The marvellous electron microscope, a by-product of modern electronics, may yet play a part comparable to that historically associated with the first optical microscope. For this we owe much to Professor Burton and his associates of the University of Toronto, where an important part of the pioneering work was done. Offering as it does useful magnifications up to say 50,000, in comparison with say 2,000 which may be taken as the limit of the ordinary microscope, it is believed to cover the last range of known magnitudes requiring investigation for medical purposes—virtually to molecular magnitudes. The light-wave microscope took care of the major range of micro-organisms and tremendously advanced the control of man over the ravages of some of them. The electron microscope will make a similar advance possible with respect to an important group of viruses and to other organisms which through their minuteness have eluded the physician and research worker. In the electron microscope we have yet another instrument developed by physicists for which the medical world had to wait before making important new advances.

In this address I have attempted to do little more than sketch the expanding role played by physics in the biological and medical fields. But enough has been said fully to support the statements of Glasser to the effect "that recent developments have led to a steadily increasing recognition of physics as one of the essential substructures of medicine, that progress in biology and in medicine has always been significantly related to advances in physics. Physics has supplied the instruments and the methods of physical research which have revolutionized many of the older approaches to the biological sciences." In spite of this there are still professors and practitioners of medicine who remain unaware of the basic importance of physics in their field, and professors of physics who teach premedical students the important role played by physics in engineering but fail to lead their students even to suspect that physics plays any important part in their sphere of interest. It is hoped that this address will in some measure lead professors of physics to give new concepts of physics to those of their students who look forward to a career in medicine, and thereby make it possible for them to use more understandingly and promptly the contributions to medicine physics has made, is making now, and may make in the future.

---

[1] Seaborg believes "artificial radioactivity has given biology and medicine what is probably the most useful tool for research since the discovery of the microscope."

UNIVERSITY OF SASKATCHEWAN

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLI : SERIES III : MAY, 1947
SECTION THREE

On a Conjecture by Nakayama

By RICHARD BRAUER, F.R.S.C.

§1. *Introduction*

As is well known, every irreducible representation of the symmetric group $\mathfrak{S}_n$ is characterized by a Young diagram. If p is a fixed prime number, the theory of modular characters furnishes a distribution of the different irreducible representations of a group into "p-blocks." The problem arises of finding a criterion which would make it possible to decide whether or not two irreducible representations of $\mathfrak{S}_n$, given by their Young diagrams, belong to the same p-block. This question is of importance for a study of the arithmetical properties of the representations of $\mathfrak{S}_n$. Some years ago, T. Nakayama[1] published a conjecjecture which offers an elegant solution of the problem. It is the purpose of this paper and a following paper by G. de B. Robinson to give a proof of Nakayama's conjecture. The proof was found jointly by Robinson and the author.

The basis for the work of this note is formed by some general results concerning the p-blocks of groups of finite order. These results which extend some earlier investigations[2] are given without proofs in section 2. The later sections contain the application to the case of the symmetric group. In this note, the number of different p-blocks of $\mathfrak{S}_n$ is determined, and it is shown that the proof of Nakayama's conjecture is equivalent to a proof of three statements concerning the characters of the symmetric groups. The proofs of these statements will be given in the paper by G. de B. Robinson. It is only there that the finer properties of the characters of the symmetric groups will be used.

§2. *The Defect Group*

Let $\mathfrak{G}$ be a group of finite order $g = p^a g'$ where p is a fixed prime number and where g' is relatively prime to p. In the following, $\mathfrak{p}$ will always denote a fixed prime ideal divisor of p in a field which contains all the characters of $\mathfrak{G}$, say in the field of the g^{th} roots of unity.

---

[1] *Jap. Jour. Math.*, **17**: 411-23 (1941).

[2] R. Brauer, *Proc. Nat. Acad. Sci. U.S.A.*, **30**: 109-14 (1944); **32**: 182-6 and 215-19 (1946).

If $\mathfrak{H}$ is any subset of $\mathfrak{G}$, the normalizer $\mathfrak{N}(\mathfrak{H})$ of $\mathfrak{H}$ is the set of all elements G of $\mathfrak{G}$ which commute with $\mathfrak{H}$, $G\mathfrak{H} = \mathfrak{H}G$. The order of $\mathfrak{N}(\mathfrak{H})$ will be denoted by $n(\mathfrak{H})$. On the other hand the centralizer $\mathfrak{C}(\mathfrak{H})$ of $\mathfrak{H}$ consists of those elements of $\mathfrak{G}$ which commute with every element of $\mathfrak{H}$. If $\mathfrak{H}$ consists of only one element H, we have $\mathfrak{N}(H) = \mathfrak{C}(H)$, and $g/n(H)$ denotes the number of elements in the class of H.

The irreducible characters of $\mathfrak{G}$ will be denoted by $\zeta_1, \zeta_2, \ldots, \zeta_k$ and we shall set

$$\omega_i(G) = g\zeta_i(G)/n(G)z_i$$

where

$$z_i = Dg(\zeta_i)$$

is the degree of ζ_i.

To every p-block B, there belongs a defect group $\mathfrak{D}$ which is uniquely determined except that it can be replaced by any conjugate group. This defect group has the following properties.[3]

I. The group $\mathfrak{D}$ is a subgroup of $\mathfrak{G}$ whose order is the highest power p^d of p dividing one of the numbers g/z_i with ζ_i in B. The exponent d here is called the defect of the block B.

II. There exist p-regular elements[4] V in $\mathfrak{G}$ such that $\mathfrak{D}$ is a p-Sylow-subgroup of $\mathfrak{N}(V)$ and that $\omega_i(V) \not\equiv 0 \pmod{\mathfrak{p}}$ for the ω_i belonging to characters ζ_i of the block B.

III. If the class of an element H of $\mathfrak{G}$ does not contain elements of the centralizer $\mathfrak{T} = \mathfrak{C}(\mathfrak{D})$, then $\omega_i(H) \equiv 0 \pmod{\mathfrak{p}}$ for the ω_i belonging to the characters ζ_i of B.

The properties I and II characterize $\mathfrak{D}$ as the defect group of B. Also, II and III can be used to characterize $\mathfrak{D}$.

IV. If ζ_i and ζ_j both belong to blocks having $\mathfrak{D}$ as their defect group and if

$$\begin{vmatrix} \omega_i(V_1) & \omega_i(V_2) \\ \omega_j(V_1) & \omega_j(V_2) \end{vmatrix} \equiv 0 \pmod{\mathfrak{p}}$$

for any pair of p-regular elements V_1 and V_2 of $\mathfrak{G}$ for which $\mathfrak{D}$ is a p-Sylow-subgroup as well of $\mathfrak{N}(V_1)$ as of $\mathfrak{N}(V_2)$, then ζ_i and ζ_j belong to the same p-block.

V. If P is an element of an order p^h, $h \geqq 1$, which is not conjugate to an element of $\mathfrak{D}$, and if V is a p-regular element commuting with P, then

$$\zeta_i(PV) = 0$$

for all ζ_i belonging to a block B with the defect group $\mathfrak{D}$.

---

[3] The first seven of these properties have been stated in the papers quoted in foot-note 2. Property VIII is obtained by a continuation of this work. Finally, IX can be derived easily from the other results. The proofs will be given in detail in a forthcoming paper.

[4] An element of a group is p-regular, if its order is prime to p.

VI. If P is an invariant element of the defect group $\mathfrak{D}$ of B, then for every ζ_i in B there exists a p-regular element V of $\mathfrak{N}(P)$ such that $\zeta_i(PV) \neq 0$. The group $\mathfrak{D}$ occurs as defect group of p-blocks of the group $\mathfrak{N}(P)$.

VII. If $\mathfrak{D}$ is not a maximal normal p-subgroup of $\mathfrak{N}(\mathfrak{D})$, then $\mathfrak{D}$ cannot occur as a defect group.

VIII. The number of p-blocks with a given defect group $\mathfrak{D}$ can be determined as follows. Set

$$\mathfrak{U} = \mathfrak{D}\mathfrak{C}(\mathfrak{D})/\mathfrak{D}, \quad \mathfrak{V} = \mathfrak{N}(\mathfrak{D})/\mathfrak{D}$$

so that $\mathfrak{U}$ is a normal subgroup of $\mathfrak{V}$. Determine all characters θ of $\mathfrak{U}$ of defect 0. Discard those θ for which there exists an element V in $\mathfrak{V}$ not in $\mathfrak{U}$ such that V^p lies in $\mathfrak{U}$ and that $\theta(V^{-1}UV) = \theta(U)$ for all U in $\mathfrak{U}$. The remaining characters θ are distributed into classes of characters associated in $\mathfrak{V}$.[5] If r is the number of such classes of characters, then $\mathfrak{G}$ possesses r p-blocks with the defect group $\mathfrak{D}$.

Actually, a (1-1) correspondence between the classes of characters and the blocks B with the defect group $\mathfrak{D}$ can be established such that the values of θ and the values of the characters ζ_i of the corresponding block B are related by congruences. However, this is not essential for our purpose.

IX. Suppose that $\mathfrak{G}$ is a direct product, $\mathfrak{G} = \mathfrak{G}_1 \times \mathfrak{G}_2$. As is well known, every character ζ of $\mathfrak{G}$ can be expressed as the product of a character ζ_1 of $\mathfrak{G}_1$ with a character ζ_2 of $\mathfrak{G}_2$. If B_1 is a p-block of $\mathfrak{G}_1$, and B_2 a p-block of $\mathfrak{G}_2$, then the $\zeta = \zeta_1\zeta_2$ with ζ_1 in B_1 and ζ_2 in B_2 will form a p-block of $\mathfrak{G}$. Every p-block B of $\mathfrak{G}$ is obtained in this form. If B_1 has the defect group $\mathfrak{D}_1$ in $\mathfrak{G}_1$, and if B_2 has the defect group $\mathfrak{D}_2$ in $\mathfrak{G}_2$, then B has the defect group $\mathfrak{D} = \mathfrak{D}_1 \times \mathfrak{D}_2$ in $\mathfrak{G}$.

§3. *Lemmas*

LEMMA 1: If the group $\mathfrak{G}$ contains a normal subgroup $\mathfrak{Q}$ whose order is a power of p, then the defect group $\mathfrak{D}$ of every p-block B contains $\mathfrak{Q}$.

PROOF. We have $\mathfrak{Q}\mathfrak{D} \geq \mathfrak{D}$. Suppose that $\mathfrak{Q}\mathfrak{D} > \mathfrak{D}$. The intersection of $\mathfrak{N}(\mathfrak{D})$ with $\mathfrak{Q}\mathfrak{D}$ contains $\mathfrak{D}$ as a proper subgroup, since the p-group $\mathfrak{Q}\mathfrak{D}$ must contain elements outside of $\mathfrak{D}$ which commute with $\mathfrak{D}$. Since $\mathfrak{N}(\mathfrak{D}) \cap \mathfrak{Q}\mathfrak{D}$ is a normal subgroup of $\mathfrak{N}(\mathfrak{D})$, it follows that $\mathfrak{D}$ is not a maximal normal p-subgroup of $\mathfrak{N}(\mathfrak{D})$. Now VII leads to a contradiction. Hence $\mathfrak{Q}\mathfrak{D} = \mathfrak{D}$, that is $\mathfrak{Q} \leq \mathfrak{D}$ as was stated.

---

[5] If θ_1 and θ_2 are two characters of a normal subgroup $\mathfrak{U}$ of a group $\mathfrak{V}$, then θ_1 and θ_2 are associated in $\mathfrak{V}$, if there exists a fixed element V_0 of $\mathfrak{V}$ such that $\theta_2(U) = \theta_1(V_0^{-1}UV_0)$ for all U in $\mathfrak{U}$.

Lemma 2: If $\mathfrak{G}$ contains a normal subgroup $\mathfrak{Q}$ whose order is a power of p and which contains its centralizer, $\mathfrak{Q} \geqq \mathfrak{C}(\mathfrak{Q})$, then $\mathfrak{G}$ possesses only one p-block. The defect group is a p-Sylow-subgroup of $\mathfrak{G}$.[6]

Proof. If $\mathfrak{D}$ is the defect group of a block B, it follows from Lemma 1 that $\mathfrak{Q} \leq \mathfrak{D}$. Hence $\mathfrak{C}(\mathfrak{D}) \leq \mathfrak{C}(\mathfrak{Q}) \leq \mathfrak{Q} \leq \mathfrak{D}$. Apply now VIII. The group $\mathfrak{U}$ here has the order 1 and θ must be the 1-character. Since every V in $\mathfrak{V}$ transforms θ into itself, $\mathfrak{V}$ must not contain any elements of order p. Then $\mathfrak{V} = \mathfrak{N}(\mathfrak{D})/\mathfrak{D}$ must have an order prime to p. If $\mathfrak{D}$ was not a p-Sylow-subgroup of $\mathfrak{G}$, it would appear as a proper subgroup of such a Sylow group $\mathfrak{P}$ and then the intersection $\mathfrak{P} \cap \mathfrak{N}(\mathfrak{D})$ would also contain $\mathfrak{D}$ as a proper subgroup. However, this is impossible when the order of $\mathfrak{N}(\mathfrak{D})/\mathfrak{D}$ is prime to p. Hence $\mathfrak{D}$ must be a Sylow group of $\mathfrak{G}$, and since we had only one θ, the lemma is established.

§4. *The Defect Groups of the p-Blocks of $\mathfrak{S}_n$*

Our next aim is to determine which subgroups of $\mathfrak{S}_n$ can appear as defect groups $\mathfrak{D}$ of p-blocks B of $\mathfrak{S}_n$. Let $\mathfrak{D}$ be such a defect group. If the defect d is not zero, the group $\mathfrak{D}$ will contain invariant elements of order p. Choose an element of this type which contains a maximal number β of cycles of length p. Since we can replace $\mathfrak{D}$ by a conjugate group, we may assume without restriction that the element in question is

$$P_\beta = (1)(2)\ldots(a)(a+1, a+2, \ldots, a+p)\ (a+p+1,$$
$$a+p+2,)\ \ldots, a+2p)\ldots(n-p+1, n-p+2, \ldots, n)$$

where

$$n = a + \beta p.$$

It follows from VI that $\mathfrak{D}$ is also the defect group of a p-block of $\mathfrak{N}(P)$. As is readily seen, the group $\mathfrak{N}(P)$ is a direct product $\mathfrak{G}_1 \times \mathfrak{G}_2$ where $\mathfrak{G}_1$ is the subgroup of $\mathfrak{S}_n$ which permutes only the first a letters and which may be identified with $\mathfrak{S}_a$. On the other hand, $\mathfrak{G}_2$ consists of those permutations of $a+1, a+2, \ldots, n$ which transform the cycles of P_β into each other. The β cycles of length p of P_β belong to $\mathfrak{G}_2$ and generate a normal subgroup $\mathfrak{Q}$ of order p^β of $\mathfrak{G}_2$. It is easily seen that $\mathfrak{Q}$ is its own centralizer in $\mathfrak{G}_2$.

---

[6] In *Ann. Math.*, **42**: 587 (1941), it was shown by C. J. Nesbitt and the author that if a group of order $g = p^a g'$ with $(p, g') = 1$ contains a normal subgroup of order p^a, then all the p-blocks are of defect a. This is a special case of Lemma 1. The question was raised whether the converse statement was true. Using Lemma 2, we can easily construct groups of order $p^a g'$, $(p, g') = 1$, which have only one p-block, necessarily of defect a, but which do not contain any normal subgroup of order p^a. A simple example is formed by $\mathfrak{S}_4$ for $p = 2$.

On account of IX, the group $\mathfrak{D}$ is the direct product of a defect group $\mathfrak{D}_1$ in $\mathfrak{S}_\alpha$ and a defect group $\mathfrak{D}_2$ in $\mathfrak{G}_2$. The subgroup $\mathfrak{Q}$ of $\mathfrak{G}_2$ just discussed satisfies the assumptions of Lemma 2. Hence $\mathfrak{D}_2$ must be a Sylow-subgroup of $\mathfrak{G}_2$. In particular, $\mathfrak{D}_2$ contains the element P_β. If the group $\mathfrak{D}_1$ had an order larger than 1, it would contain invariant elements P' of order p. Then $P'P_\beta$ would be an invariant element of $\mathfrak{D} = \mathfrak{D}_1\mathfrak{D}_2$ of order p and $P'P_\beta$ would contain more than β cycles of length p. This is impossible from the manner in which P_β was selected. Hence $\mathfrak{D}_1 = 1$, $\mathfrak{D} = \mathfrak{D}_2$. We thus have proved the following theorem:

THEOREM 1: The defect group of a p-block B of $\mathfrak{S}_n$ can be obtained by the following construction. Determine non-negative integers α, β such that $n = \alpha + \beta p$. Consider the group of all permutations of $\alpha + 1, \alpha + 2, \dots, n$ which transform each of the cycles $Z_1 = (\alpha + 1, \alpha + 2, \dots, \alpha + p)$, $Z_2 = (\alpha + p + 1, \alpha + p + 2, \dots, \alpha + 2p), \dots,$ $Z_\beta = (n - p + 1, n - p + 2, \dots, n)$ into a cycle of the same system. If $\mathfrak{D}^{(\beta)}$ denotes a p-Sylow-subgroup of this group of permutations of degree $n - \alpha$, the defect group $\mathfrak{D}$ of B is conjugate to the group $\mathfrak{D}^{(\beta)}$ for a suitable β, $0 \leqq \beta \leqq [n/p]$.

The case $\mathfrak{D} = 1$ corresponds to the case $\beta = 0$. Here, $\mathfrak{D}^{(0)}$ is to be interpreted as a group of order 1.

It is easy to determine the defect of a block B with the defect group $\mathfrak{D}^{(\beta)}$. According to I, this defect d_β is the exponent of the highest power of p which divides the order of $\mathfrak{D}^{(\beta)}$. If $e(m)$ denotes the order with which p divides the integer m, then it is readily seen that

$$d_\beta = \beta + e(\beta!).$$

We notice that d_β increases with β.

§5. *The Number of Blocks of $\mathfrak{S}_n$ with a Given Defect Group*

THEOREM 2: The number of blocks of $\mathfrak{S}_n$ with a given defect group $\mathfrak{D}^{(\beta)}$, $n = \alpha + p\beta$, is equal to the number of irreducible characters of $\mathfrak{S}_\alpha$ whose degree is divisible by the full power of p dividing $\alpha!$.

PROOF: The cycles $Z_1, Z_2, \dots, Z_\beta$ in Theorem 1 and their powers are the only cycles of length p in $\mathfrak{D}^{(\beta)}$. Every element of $\mathfrak{N}(\mathfrak{D}^{(\beta)})$ must transform a cycle of length p of $\mathfrak{D}^{(\beta)}$ into a cycle of the same kind. It follows that $\mathfrak{N}(\mathfrak{D}^{(\beta)})$ is a direct product $\mathfrak{S}_\alpha \times \mathfrak{W}$ of $\mathfrak{S}_\alpha$ and a group of transformations $\mathfrak{W}$ of $\alpha + 1, \alpha + 2, \dots, n$. Every element W of $\mathfrak{W}$

effects a permutation $\pi(W)$ of the β cyclic subgroups generated by $Z_1, Z_2, \ldots, Z_\beta$, respectively. Let $\mathfrak{W}_1$ be a p-Sylow-subgroup of $\mathfrak{W}$ and let $\mathfrak{W}_2$ be the subgroup consisting of the elements W of $\mathfrak{W}_2$ with $\pi(W) = 1$. The group generated by an element W_2 of $\mathfrak{W}_2$ and by a cycle Z_i is abelian. Therefore, W_2 transforms Z_i into itself. This implies that W_2 lies in the group $\mathfrak{Q}$ (in the notation of section 4). Hence $\mathfrak{W}_2 \leq \mathfrak{Q}$. Since $\mathfrak{W}_1/\mathfrak{W}_2$ possesses a faithful representation by permutations of degree β, the order of $\mathfrak{W}_1/\mathfrak{W}_2$ is a divisor of $p^{e(\beta!)}$ and the order of $\mathfrak{W}_1$ then is a divisor of $p^{\beta + e(\beta!)}$. But this is the order of $\mathfrak{D}^{(\beta)}$ and since the Sylow group $\mathfrak{W}_1$ of $\mathfrak{W}$ contains $\mathfrak{D}^{(\beta)}$, it follows that $\mathfrak{W}_1 = \mathfrak{D}^{(\beta)}$. Consequently, $\mathfrak{W}/\mathfrak{D}^{(\beta)}$ has an order prime to p.

Apply now VIII. We have $\mathfrak{C}(\mathfrak{D}^{(\beta)}) \leq \mathfrak{C}(\mathfrak{Q}) = \mathfrak{S}_a \times \mathfrak{Q}$, $\mathfrak{D}^{(\beta)}\mathfrak{C}(\mathfrak{D}^{(\beta)}) = \mathfrak{S}_a \times \mathfrak{D}^{(\beta)}$, $\mathfrak{U} = \mathfrak{S}_a$ and $\mathfrak{V} = \mathfrak{S}_a \times (\mathfrak{W}/\mathfrak{D}^{(\beta)})$. No two different characters of $\mathfrak{U}$ are associated in $\mathfrak{V}$. On the other hand, if the p^{th} power of an element V of $\mathfrak{V}$ lies in $\mathfrak{U}$, then V lies in $\mathfrak{U}$. Observing that every block of defect 0 consists of only one character, we obtain Theorem 2.

Nakayama determined the characters of $\mathfrak{S}_a$ whose degree is divisible by the full power $p^{e(a!)}$. He showed that they belong to the Young diagrams with a nodes which do not contain any p-hooks.[7] If we start from an arbitrary Young diagram with n nodes and remove p-hooks as long as possible, we finally arrive at a diagram without p-hooks. We shall call this new diagram the *p-core* of the given diagram. The number of nodes of this p-core is an integer a, $0 \leqq a \leqq n$, with $n \equiv a \pmod{p}$. We can then find β such that $n = a + p\beta$. As was shown by Nakayama, the p-core is uniquely determined by the given diagram.

We have the following necessary conditions for the p-core. It is a Young diagram with a nodes, $0 \leqq a \leqq n$, $a \equiv n \pmod{p}$, such that no p-hook can be removed. It can be shown without difficulty that every such diagram actually appears as the p-core of diagrams with n nodes. The following proposition is now a consequence of Theorem 2.

PROPOSITION 1: The number of p-blocks of $\mathfrak{S}_n$ is equal to the number of possible p-cores of Young diagrams with n nodes.

---

[7] A hook in a Young diagram consists of a fixed node X, all the nodes underneath, and all the nodes to the right of X. A p-hook is a hook which consists of exactly p nodes. If a p-hook is removed from a Young diagram with n nodes, a Young diagram with n-p nodes is obtained.

§6. *Reduction of Nakayama's Conjecture*

Our Proposition 1 is part of Nakayama's conjecture. In full, Nakayama's conjecture states that two irreducible characters of $\mathfrak{S}_n$ belong to the same p-block, if and only if their Young diagrams have the same p-core.

We show here

Proposition 2: In order to prove Nakayama's conjecture, it is sufficient to prove the following three statements:

(A) Let ζ_ι be a character of $\mathfrak{S}_n$ whose Young diagram has a p-core containing α nodes, $n = \alpha + p\beta$. If $\delta > \beta$, $n = \gamma + p\delta$, and

$$P_\delta = (\gamma + 1, \gamma + 2, \ldots, \gamma + p)(\gamma + p + 1, \gamma + p + 2, \ldots, \gamma + 2p)\ldots(n - p + 1, n - p + 2, \ldots, n),$$

then $\zeta_\iota\,(V_0 P_\delta) = 0$ for all p-regular elements V_0 of $\mathfrak{S}_\gamma$.

(B) If ζ_ι satisfies the same assumption as in (A), there exists a p-regular element V_0^1 in $\mathfrak{S}_\alpha$ such that $\zeta_\iota(V_0^1 P_\beta) \neq 0$.

(C) Suppose that the character ζ_ι of $\mathfrak{S}_n$ belongs to a p-block with the defect group $\mathfrak{D}^{(\beta)}$. Suppose that g/z_ι contains p to the exponent $d_\beta - \epsilon$ so that $\epsilon \geqq 0$ by I. Then

$$\zeta_\iota(VP_\beta) \equiv k(\zeta_\iota)p^\epsilon\theta(V) \qquad (\text{mod } p^{\epsilon+1})$$

for all p-regular elements V of $\mathfrak{S}_\alpha$ whose normalizer in $\mathfrak{S}_\alpha$ has an order prime to p. Here, $k(\zeta_\iota)$ is an integer depending on ζ_ι but not on V, and $\theta(V)$ is an integer depending on V and the p-core of the Young diagram of ζ_ι but not on the diagram itself.

Proof: 1. We shall show that (A) implies that ζ_ι belongs to a block of defect $d \leqq d_\beta$. Assume that this is not so. Then ζ_ι must belong to a block B whose defect group is a group $\mathfrak{D}^{(\delta)}$, $\delta > \beta$. By II, there exists a p-regular element V_0 such that $\mathfrak{D}^{(\delta)}$ is a Sylow group of $\mathfrak{N}(V_0)$ and that $\omega_\iota(V_0) \not\equiv 0 \pmod{\mathfrak{p}}$. Then V_0 can permute only the first $\gamma = n - p\delta$ letters. Choose a character ζ_h in B whose degree is divisible by a minimal power of p. It follows from I that $g\ z_h$ contains p with the order d_δ. Since ζ_ι and ζ_h belong to the same block, $\omega_\iota(V_0) \equiv \omega_h(V_0) \pmod{\mathfrak{p}}$. Expressing ω_h by means of ζ_h and using the properties of z_h and V_0, we derive $\zeta_h(V_0) \not\equiv 0 \pmod{\mathfrak{p}}$. But $\zeta_h(V_0) \equiv \zeta_h(P_\delta V_0) \pmod{\mathfrak{p}}$. Taking into account that $\mathfrak{D}^{(\delta)}$ is a Sylow group of $\mathfrak{N}(P_\delta V_0)$ and introducing again ω_h, we obtain $\omega_h(P_\delta V_0) \not\equiv 0 \pmod{\mathfrak{p}}$. This implies $\omega_\iota(P_\delta V_0) \not\equiv 0 \pmod{\mathfrak{p}}$, and then, certainly, $\zeta_\iota(P_\delta V_0) \neq 0$. However, this contradicts (A) and ζ_ι must belong to a block B of defect $d \leqq d_\beta$.

2. Assume now that (B) is true and that ζ_i belongs to a block B of defect $d < d_\beta$. Then the defect group of B is a group $\mathfrak{D}^{(\delta)}$ with $\delta < \beta$. The group $\mathfrak{D}^{(\delta)}$ does not contain elements consisting of more than δ cycles of length p. In particular, P_β is not conjugate to an element of $\mathfrak{D}^{(\delta)}$. By V, $\zeta_i(P_\beta V_0) = 0$ for all p-regular elements V_0 of $\mathfrak{S}_\alpha$. This contradicts (B). Hence ζ_i must belong to a block of defect d_β. The defect group then is $\mathfrak{D}^{(\beta)}$.

3. We have shown in 1 and 2 that if the Young diagram of ζ_i has a p-core with α nodes, then ζ_i belongs to a block with the defect group $\mathfrak{D}^{(\beta)}$ where $n = \alpha + p\beta$. Let V be a p-regular element such that $\mathfrak{D}^{(\beta)}$ is a Sylow group of $\mathfrak{N}(V)$. Then V must belong to $\mathfrak{S}_\alpha$ and the normalizer of V in S_α has an order prime to p. Conversely, if V is an element of $\mathfrak{S}_\alpha$ whose normalizer in $\mathfrak{S}_\alpha$ has an order prime to p, the elements of a Sylow group $\mathfrak{D}'$ of $\mathfrak{N}(V)$ will leave at least $\alpha - p + 1$ of the letters $1, 2, \ldots, \alpha$ invariant. Hence $\mathfrak{D}'$ has at most the order p^r with $r \leqq e(p - 1 + p\beta!) = e(p\beta!) = d_\beta$. It follows that $\mathfrak{D}^{(\beta)}$ is a Sylow group of $\mathfrak{N}(V)$. Then $\mathfrak{D}^{(\beta)}$ is also a Sylow group of $\mathfrak{N}(VP_\beta)$, since P_β belongs to the center of $\mathfrak{D}^{(\beta)}$. If we set $d_\beta = d$, we can write

$$n(VP_\beta) = p^d n'(VP_\beta),\ n(V) = p^d n'(V)$$

where $n'(VP_\beta)$ and $n'(V)$ are integers which are prime to p. Set

$$z_i = p^{a-d+\epsilon} z_i'$$

with an integral z_i' relatively prime to p. If as was assumed the Young diagram of ζ_i has a p-core with α nodes, then $\epsilon \geqq 0$ by I. Now

$$\omega_i(VP_\beta) = g'\zeta_i(VP_\beta)/n'(VP_\beta)p^\epsilon z_i'.$$

Applying (C), we find

$$(1) \qquad \omega_i(VP_\beta) \equiv g'k(\zeta_i)\theta(V)/n'(VP_\beta)z_i' \qquad (\text{mod } \mathfrak{p}).$$

Since the block B containing ζ_i has the defect d, B will contain a character ζ_h for which g/z_h contains p with the exact order d. From the fundamental properties of p-blocks, it follows that

$$(2) \qquad \omega_i(VP_\beta) \equiv \omega_h(VP_\beta), \quad \omega_i(V) \equiv \omega_h(V) \qquad (\text{mod } \mathfrak{p}).$$

On the other hand, we have

$$(3) \qquad \zeta_h(VP_\beta) \equiv \zeta_h(V) \qquad (\text{mod } \mathfrak{p}).$$

Express ζ_h by ω_h in (3). This gives

$$n(VP_\beta)z_h\omega_h(VP_\beta)/g \equiv n(V)z_h\omega_h(V)/g \qquad (\text{mod } \mathfrak{p})$$

from which

$$\omega_h(V) \equiv n'(VP_\beta)\omega_h(VP_\beta)/n'(V) \qquad (\text{mod } \mathfrak{p})$$

can be derived without difficulty. Combining this with (1) and (2), we find

$$\omega_i(\Gamma) \equiv g'k(\zeta_i)\theta(\Gamma)/n'(\Gamma)z_i' \pmod{\mathfrak{p}}. \tag{4}$$

Let ζ_j be a second character whose Young diagram has the same p-core as that of ζ_i. Then ζ_j also belongs to a block with the defect group $\mathfrak{D}^{(\beta)}$. The expression $\theta(\Gamma)$ is the same for ζ_i and ζ_j. Applying (4) to the two characters ζ_i and ζ_j, we see that

$$\begin{matrix} \omega_i(\Gamma_1) & \omega_i(\Gamma_2) \\ \omega_j(\Gamma_1) & \omega_j(\Gamma_2) \end{matrix} \qquad 0 \qquad (\bmod\ \mathfrak{p})$$

for any two p-regular elements Γ_1 and Γ_2 for which $\mathfrak{D}^{(\beta)}$ is a Sylow group of $\mathfrak{N}(\Gamma_1)$ and of $\mathfrak{N}(\Gamma_2)$. Now IV shows that ζ_i and ζ_j belong to the same p-block. Assuming the statements (A), (B), (C) we have thus shown that if the Young diagrams of two characters have the same p-core, both characters belong to the same p-block. According to Proposition 1, the number of p-blocks is equal to the number of possible p-cores. Hence if two characters belong to the same p-block, their Young diagrams must necessarily have the same p-core. This shows that Nakayama's conjecture is correct, if (A), (B), (C) are assumed.

It remains to show that (A), (B), (C) are actually true. This proof will be given by G. de B. Robinson in the following paper.

UNIVERSITY OF TORONTO

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLI : SERIES III : MAY, 1947
SECTION THREE

On a Conjecture by Nakayama

By G. de B. ROBINSON, F.R.S.C.

§1. *Introduction*

This paper is a sequel to one bearing the same title by Richard Brauer which will be referred to here as RB. Taken together, the two papers provide a proof of a conjecture by Nakayama concerning the blocks of characters of the symmetric group $\mathfrak{S}_n$. The problem is clearly stated in RB and it remains to prove the statements (A), (B), (C) of Proposition 2.

The basic requirement is to provide an operational approach to Nakayama's "removal of a hook" from a Young diagram. This necessitates the generalization of the familiar *right* diagram $[\lambda]$ to the *skew* diagram $[\lambda] - [\mu]$, consisting of those nodes of $[\lambda]$ not belonging to $[\mu]$. The symbol has a meaning if and only if the right diagram $[\mu]$ is contained in $[\lambda]$ when the top left-hand corner nodes of the two diagrams are brought into coincidence.[1] The necessary formulae involving the characters will be given in the following section and an attempt will be made to suggest the underlying argument. For further details the reader is referred to SG$_2$ and SG$_3$. These formulae, as they stand, yield statements (A) and (B).

To prove (C) it is necessary to refine Nakayama's theorem concerning the prime factors of the degree of an irreducible representation of $\mathfrak{S}_n$. The necessary theorem is given in section 3.

The notation used will conform as far as possible with that of RB. As is customary, the symbol $[\lambda]$ will denote a specific Young diagram, where $n = \lambda_1 + \lambda_2 + \ldots$ and $\lambda_i \geq \lambda_{i+1}$. It will sometimes be desirable to make the n explicit by writing $\lambda(n)$. The degree of the irreducible representation $[\lambda]$, with character ζ_λ, will be written f_λ, while the degree of the reducible representation $[\lambda] - [\mu]$, with character ζ^{μ}_{λ}, will be written z^{μ}_{λ}.

---

[1] G. de B. Robinson, "On the representations of the symmetric group," *Amer. Jour. Math.*, **60**: 745-60 (1938); Second Paper, *ibid.*, **69**: 286-98 (1947). Third Paper to appear shortly. These papers will be referred to as SG$_1$, SG$_2$, SG$_3$. The reference here is to SG$_2$, section 2.

Since the appearance of SG$_2$ the author's attention has been drawn to a paper by A. C. Aitken, *Proc. Roy. Soc. of Edinburgh*, (A) **61**: 300-10 (1943). Here the notion of a skew diagram is introduced and the theorem giving the reduction of the corresponding reducible representation is proved, though the language used is that of Schur-functions. Further reference to this paper will be made in SG$_3$.

§2. Nakayama recognized the significance of the *skew* hook consisting of the nodes on the rim of a diagram determined by a *right* hook, but made no attempt to develop the idea of the corresponding skew representation. For example, if $[\lambda] = [4, 3, 2]$, $[\mu] = [2, 1]$ we have the diagram

$$(2.1) \qquad \begin{array}{cccc} \circ & \circ & \cdot & \cdot \\ \circ & \cdot & \cdot & \\ \cdot & \cdot & & \end{array},$$

and the irreducible components of the skew representation are given by[2]

$$(2.2) \qquad [\lambda] - [\mu] = [4, 2] + [4, 1^2] + [3^2] + 2[3, 2, 1] + [2^3],$$

with $z^{\mu}_{\lambda} = 61$. The skew hook $[\lambda] - [\mu]$ is *equivalent* to the right hook $H = [4, 1^2]$ in the sense that each determines the other. It was proved in SG$_2$ that every skew hook representation contains the equivalent right hook representation exactly once as an irreducible component, and that[3]

$$(2.3) \qquad \zeta^{\mu(m)}_{\lambda(n)}\,(C) = (-1)^r \text{ or } 0,$$

where C is a cycle on $q = n - m$ symbols, according as the skew diagram $[\lambda] - [\mu]$ is or is not a skew hook $H = [q-r, 1^r]$.

Instead of considering the matrices in $[\lambda]$ associated with all the substitutions of $\mathfrak{S}_n$ let us restrict attention to those associated with the substitutions of the direct product $\mathfrak{S}_m \times \mathfrak{S}_q$. These matrices generate a representation of $\mathfrak{S}_m \times \mathfrak{S}_q$ which is reducible and for these substitutions we have the relation[4]

$$(2.4) \qquad \zeta_{\lambda(n)} = \sum_{\mu} \zeta_{\mu(m)} \cdot \zeta^{\mu(m)}_{\lambda(n)}.$$

between the characters. An irreducible representation of $\mathfrak{S}_m \times \mathfrak{S}_q$ is necessarily the Kronecker product of an irreducible representation of $\mathfrak{S}_m$ and an irreducible representation of $\mathfrak{S}_q$, while the character of such a representation is the product of the corresponding characters. The terms on the right-hand side of (2.4) are not irreducible but yield the irreducible components when ζ^{μ}_{λ} is replaced by a sum of irreducible characters, as in (2.2).

Since there is only one hook of length 6 in $[\lambda] = [4, 3, 2]$ there is only one term in the appropriate form of (2.4) which corresponds to the removal of a skew hook; actually, there are three terms in the summation, the other two corresponding to $[\mu] = [3]$ and $[1^3]$.

---

[2] SG$_2$, section 3.

[3] This generalizes a result of Murnaghan, cf. SG$_3$, sections 6, 7.

[4] Cf. RB IX.

Consider the successive removal of β hooks of length p, where p is assumed to be prime and β is as large as possible. Either all the nodes of $[\lambda(n)]$ are exhausted or a *p-core* $[\lambda_0(\alpha)]$ remains from which no further p-hooks can be removed, and $n = \alpha + \beta p$. Iteration of (2.4) yields the equation

$$(2.5) \qquad \zeta_{\lambda(n)} = \Sigma \zeta_{\mu(\alpha)} \underbrace{\zeta_{(p)} \zeta_{(p)} \cdots \zeta_{(p)}}_{\beta \text{ factors}},$$

which is applicable to the substitutions of $\mathfrak{S}_n$ belonging to the direct product of $\mathfrak{S}_\alpha$ and β factors $\mathfrak{S}_p$; $\zeta_{(p)}$ is the character of the appropriate substitution of $\mathfrak{S}_p$ in a skew representation whose corresponding skew diagram has p nodes. If we let P_β be the product of β cycles of length p on the last βp of the n symbols, as in RB, then it follows from (2.3) that the only terms which contribute to the summation in (2.5) are those arising from terms in which *every* factor corresponds to a p-hook. If z'_p is the number of ways of removing the β hooks of length p, then

$$(2.6) \qquad \zeta_{\lambda(n)}(\Gamma P_\beta) = \sigma \,.\, z'_p \,.\, \zeta_{\lambda_0(\alpha)}(\Gamma),$$

where $\sigma = \pm 1$ depending on the product of the parities (2.3); this product of parities is invariant under different orders of removal of the p-hooks, and so depends only on $[\lambda]$. Statement (B) follows from (2.6), since $[\lambda_0(\alpha)]$ is a representation of $\mathfrak{S}_\alpha$ whose degree z_{λ_0} is divisible by the maximum power of p contained in $\alpha!$ (cf. section 3) and so constitutes by itself a block of highest kind. The character $\zeta_{\lambda_0}(\Gamma)$ vanishes for all p-singular elements of $\mathfrak{S}_\alpha$, so there must be at least one p-regular element Γ_0 for which $\zeta_{\lambda_0}(\Gamma_0) \neq 0$.

If δ skew diagrams, each consisting of p nodes, where $\delta > \beta$, are removed from $[\lambda]$, then not all of them can be skew hooks since we have assumed that $[\lambda_0(\alpha)]$ is the p-core of $[\lambda]$. Hence it follows from (2.3) that

$$(2.7) \qquad \zeta_{\lambda(n)}(\Gamma P_\delta) = 0,$$

for *every* Γ of $\mathfrak{S}_\gamma$, where $n = \gamma + \delta p$. Certainly, then, this relation holds for the p-regular elements Γ_0 of $\mathfrak{S}_\gamma$, which proves statement (A).

We may generalize (2.6) by considering those substitutions of $\mathfrak{S}_n$ which permute the β cycles of P_β amongst themselves.[5] The normalizer $\mathfrak{N}(P_\beta)$ of P_β is of the form

$$\mathfrak{N}(P_\beta) \simeq \mathfrak{S}_\alpha \times \mathfrak{P} \times \mathfrak{S}'_\beta,$$

---

[5] Cf. RB, section 4.

where $\mathfrak{P}$ is the group generated by the β cycles of length p, and $\mathfrak{S}_\beta$ is isomorphic to the group generated by those substitutions W of $\mathfrak{S}_n$ which permute the cycles of P_β and which leave each of the first a symbols unaltered. We obtain a representation of $\mathfrak{S}_\beta$ by considering the p-hook structure of $[\lambda]$. This structure leads to a skew diagram $[\lambda]_p$ such that every node of $[\lambda]'_p$ represents a p-hook of $[\lambda]$ and every r-hook of $[\lambda]_p$ represents an rp-hook of $[\lambda]$, and conversely. The degree of $[\lambda]_p$ is just z'_p. Thus it is natural that we should have

$$\zeta_{\lambda(n)}(VP_\beta W) = \sigma \, . \, \zeta'_p(W^*) \, . \, \zeta_{\lambda_0(a)}(V), \tag{2.8}$$

where W^* is the image of W in $\mathfrak{S}'_\beta$. Analogous to (2.7) we have

$$\zeta_\lambda(VP_\delta W) = 0, \tag{2.9}$$

for every V of $\mathfrak{S}_\gamma$ and every W which permutes the cycles of P_δ but leaves each of the first γ symbols unaltered.

The skew diagram $[\lambda]'_p$ is of some interest and its form will be of importance in the following section. To have a clearer picture of it consider the following simple example. Take $n = 12$, $[\lambda] = [7, 2, 1^3]$, $p = 2$, so that[6]

$[\lambda]_2$

as may be easily verified. Evidently $\sigma = -1$, and the three 2-hooks of $[\lambda]$ can be removed in three different ways; this is just the number of ways in which nodes can be removed from $[\lambda]'_2$, or the number of *standard* skew diagrams $[\lambda]'_2$ which is the degree z'_2.

In general $[\lambda]'_p$ is a skew diagram which is made up of at most p right diagrams which are *disjoint*, i.e. no two have a row or column in common. Evidently, the number of ways in which the symbols can be arranged in such a skew diagram so that they follow the natural order in both rows and columns is given by

$$z'_p = \frac{\beta!}{\beta_1!\,\beta_2!\ldots\beta_p!}\, z_{\beta_1} \cdot z_{\beta_2} \cdots z_{\beta_p}, \tag{2.10}$$

where $\beta = \beta_1 + \beta_2 + \ldots + \beta_p$, and z_{β_1} is the degree of the right constituent $[\beta_1]$, etc. We may call $[\lambda]'_p$ the *disjoint* product representation of $\mathfrak{S}'_\beta$ which is induced by the Kronecker product representation

$$[\beta_1] \times [\beta_2] \times \ldots \times [\beta_p]$$

---

[6] The details of the construction of $[\lambda]^*_p$ are given in SG3, section 4.

of the sub-group $\mathfrak{H}$, where $\mathfrak{H}$ is the direct product

$$\mathfrak{S}_{\beta_1} \times \mathfrak{S}_{\beta} \times \ldots \times \mathfrak{S}_{\beta_p}.$$

§3. Instead of removing the β hooks of length p individually, Nakayama[7] thought of removing hooks whose lengths are powers of p; such a p^r-hook would be equivalent to p^{r-1} distinct p-hooks. He introduced the notion of the *p-series* of $[\lambda]$ to have the following meaning: take off first the longest p^r-hook, of length say p^{e_1}, then the longest p^r-hook in the residual diagram whose length does not exceed p^{e_1}, say p^{e_2}, and so on, until all the p-hooks have been removed from $[\lambda]$. The p-series of $[\lambda]$ would then be

$$p^{e_1},\ p^{e_2},\ \ldots, \tag{3.1}$$

where $e_1 \geq e_2 \geq \ldots.$ Nakayama proved that

$$e(z_\lambda) = e(n!) - \sum_i e(p^{e_i}!), \tag{3.2}$$

where $e(z)$ is the exponent of p dividing z.

We refine (3.2) by introducing the representation $[\lambda]_p^{\prime}$ on β nodes, where $n = \alpha + \beta p$, and prove

THEOREM 1. $e(z_\lambda) = e(n!) - e((n - \alpha)!) + e(z_p^{\prime}).$

Consider first the proof in the case that $\alpha = 0$, i.e. that $[\lambda]$ has no p-core. By (3.2)

$$\begin{aligned} e(z_\lambda) &= \left\{\left[\frac{n}{p}\right] + \left[\frac{n}{p^2}\right] + \ldots\right\} - \sum_i (p^{e_i-1} + p^{e_i-2} + \ldots) \\ &= \beta + e(\beta!) - \sum p^{e_i-1} - \sum e(p^{e_i-1}!). \end{aligned} \tag{3.3}$$

But each p^{e_i}-hook of $[\lambda]$ is represented by a p^{e_i-1}-hook of $[\lambda]_p^{\prime}$, so that $\beta = \sum_i p^{e_i-1}$ and the theorem follows immediately from (3.2) provided $[\lambda]_p^{\prime}$ consists of a single right diagram. If $[\lambda]_p^*$ has more than one constituent, it follows from (2.10) that

$$e(z_p^*) = e(\beta!) - \sum_{i=1}^{p} [e(\beta_i!) - e(z_{\beta_i})], \tag{3.4}$$

in which the summation on the right is over the hooks of length p^{e_i-1} of $[\lambda]_p^*$, by (3.2). Every such hook must belong to a unique constituent of $[\lambda]_p^*$, so that $e(z_\lambda) = e(z_p^*)$.

---

[7] *Jap. Jour. Math.*, **17**: 165-84 (1940), sections 7, 8.

If $[\lambda]$ has a p-core $[\lambda_0]$ and $a > 0$, the argument of the simpler case is applicable, after writing 3.2 in the form

$$
\begin{aligned}
(3.5)\qquad e(z_\lambda) &= e(n!) - e((n-a)!) + e((n-a)!) - \sum_i e(p^{e_i}!) \\
&= e(n!) - e((n-a)!) + e(z_p).
\end{aligned}
$$

Since the p-series of $[\lambda]$ does not enter explicitly into the statement of Theorem 1, one would expect that the notion could be dispensed with and the whole emphasis laid on the skew diagram $[\lambda]_p^{\prime}$ which completely describes the p-hook structure of $[\lambda]$.

We can rewrite Theorem 1 in the form

$$e(n!) - e(z_\lambda) = e((n-a)!) - e(z_p),$$

and this is just the power of p dividing g/z_λ, where $g = n!$. Now it was shown in RB that

$$
\begin{aligned}
d_\beta &= \beta + e(\beta!) = e((p\beta)!) \\
&= e((n-a)!),
\end{aligned}
$$

which proves

THEOREM 2. For the symmetric group $\mathfrak{S}_n$,

$$\epsilon = e(z_p) \geq 0,$$

since z_p^* is just the number of ways of removing the p-hooks of $[\lambda]$.

We obtain statement (C) by combining Theorem 2 with 2.6. Actually, our result is stronger than is required since we have an equality instead of a congruence. Brauer's $\theta(\Gamma)$ can be taken to be the character $\zeta_{\lambda_0}(\Gamma)$, and $(k\zeta_i) = \sigma\kappa$ where

$$z_p^* = p^{\epsilon}(\kappa + \kappa_1 p + \dots).$$

UNIVERSITY OF TORONTO

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLI : SERIES III : MAY, 1947
SECTION THREE

Oscillating Satellites with the Force Varying Inversely as the n^{th} Power of the Distance

By DANIEL BUCHANAN, F.R.S.C.

§ 1. *The Problem Considered*

This paper deals with the libration points and with the periodic orbits in their vicinity in the restricted problem of three bodies when the law of attraction is assumed to vary inversely as the n^{th} power of the distance, n being any real positive rational number.

In the restricted problem of three bodies, two of the bodies are assumed to be finite and to move in circles about their centre of gravity. Their masses are denoted by $1 - \mu$ and μ and the notation is so chosen that $\mu \leqq 1/2$. The third body is subject to their attraction but it is infinitesimal, that is, it is assumed to be so small relative to the other bodies that the influence of its attraction upon them may be neglected.

In 1772 Lagrange found the five libration points for the Newtonian law, three called the straight-line and two the equilateral-triangle points. At a point of libration or an equilibrium point, the centripetal force of attraction equals the centrifugal force of revolution. The straight-line points are stable, that is, periodic orbits exist in their vicinity for all values of $\mu \leqq 1/2$ but the equilateral-triangle points are stable only if $\mu \angle .0385 \ldots$.

Periodic orbits of two and of three dimensions near the libration points, called Oscillating Satellites, were shown to exist and were constructed by Moulton (1920) for the straight-line points and by Buck (1920) for the equilateral-triangle points—both for the Newtonian law.

In the present paper the same five equilibrium points are found for the general value of n, and two- and three-dimensional oscillating satellites are constructed for each set of libration points. There is no restriction upon n or μ for the oscillating satellites of either two or three dimensions near the straight-line points. For the equilateral-triangle points two-dimensional oscillating satellites will exist only if $n \angle 3$ and further, if n lies between 1.144 and 3, a restriction depending upon n must be placed upon μ. This restriction reduces to $\mu \angle .0385\ldots$ for $n = 2$. In the case of the three-dimensional oscillating satellites near the equilateral-triangle points there is no restriction upon μ but certain possible values of n in the range $0 \angle n \angle 3$ are excluded which will render certain characteristic exponents commensurable with $\sqrt{-1}$.

No attempt is made to construct orbits corresponding to Class C of Moulton's (1920) Oscillating Satellites, that is, three-dimensional orbits whose period is a multiple of the periods of the two- and three-dimensional orbits.

§ 2. *The Differential Equations of Motion*

We shall adopt the same notation, units, and system of co-ordinates as used in Moulton's *Celestial Mechanics* (1914). The finite bodies are assumed to move uniformly in circles about their centre of mass, and this point is taken as the origin of the co-ordinate system presently defined.

A system of rotating co-ordinate axes $\xi\eta\zeta$ is taken having the $\xi\eta$-plane as co-incident with the plane of revolution of the finite bodies. The rate of rotation is chosen so that the ξ-axis always passes through the centres of the finite bodies.

The units of length, mass, and time, respectively, are chosen so that the distance between the centres of the finite bodies, the sum of the masses, and the Gaussian constant are all unity. With the units thus chosen the angular velocity of revolution is likewise unity.

The masses are denoted by $1 - \mu$ and μ, $0 \angle \mu \leqq 1/2$, and their co-ordinates by $(\xi_1, 0, 0)$ and $(\xi_2, 0, 0)$, respectively, where

$$\xi_1 = -\mu, \quad \xi_2 = 1 - \mu.$$

If the force of attraction is assumed to vary inversely as the n^{th} power of the distance, the differential equations of motion of the infinitesimal body are

$$\left.\begin{aligned}
\frac{d^2\xi}{dt^2} - 2\frac{d\eta}{dt} &= \frac{\partial U}{\partial \xi} = \xi - (1-\mu)\frac{\xi - \xi_1}{r_1^{n+1}} - \mu\frac{\xi - \xi_2}{r_2^{n+1}}, \\
\frac{d^2\eta}{dt^2} + 2\frac{d\xi}{dt} &= \frac{\partial U}{\partial \eta} = \eta - (1-\mu)\frac{\eta}{r_1^{n+1}} - \mu\frac{\eta}{r_2^{n+1}}, \\
\frac{d^2\zeta}{dt^2} &= \frac{\partial U}{\partial \zeta} = - (1-\mu)\frac{\zeta}{r_1^{n+1}} - \mu\frac{\zeta}{r_2^{n+1}},
\end{aligned}\right\} \qquad (1)$$

$$U = \tfrac{1}{2}(\xi^2 + \eta^2) + \frac{1-\mu}{(n-1)r_1^{n-1}} + \frac{\mu}{(n-1)r_2^{n-1}}, \quad n \angle 1,$$

$$= \tfrac{1}{2}(1-\mu)\left[r_1^2 + \frac{2}{(n-1)r_1^{n-1}}\right] + \tfrac{1}{2}\mu\left[r_2^2 + \frac{2}{(n-1)r_2^{n-1}}\right]$$

$$- \tfrac{1}{2}\zeta^2 - \tfrac{1}{2}\mu(1-\mu),$$

$$U = \tfrac{1}{2}\,(\xi^2 + \eta^2) + (1 - \mu)\log r_1 + \mu \log r_2,\ n = 1,$$
$$r_1 = [(\xi + \mu)^2 + \eta^2 + \zeta^2]^{1/2},$$
$$r_2 = [(\xi - 1 + \mu)^2 + \eta^2 + \zeta^2]^{1/2}.$$

PART I

THE EQUILIBRIUM POINTS

§ 3. *The Straight-Line Equilibrium Points*

The straight-line equilibrium points are found by putting $\eta = \zeta = 0$ and solving

$$F(\xi) \equiv \xi - \frac{(1 - \mu)\,(\xi - \xi_1)}{[(\xi - \xi_1)^2]^{(n+1)/2}} - \frac{\mu(\xi - \xi_2)}{[(\xi - \xi_2)^2]^{(n+1)/2}} = 0. \qquad (2)$$

To show the existence of equilibrium points it is necessary that $F(\xi)$ should have poles at $\xi = \xi_1$ and $\xi = \xi_2$. This will be the case only if n is positive. Hence we rule out the case for negative values of n. If n is a rational fraction l/k, where l and k are relatively prime integers, we shall agree to take the real k^{th} root, if k is odd, or the real positive k^{th} root if k is even. With these arrangements concerning n it will be found that $F(\xi)$ is single-valued, finite, and continuous, except at $\xi = \xi_1$ and $\xi = \xi_2$ where it has poles. Further $F(\xi)$ is posi-

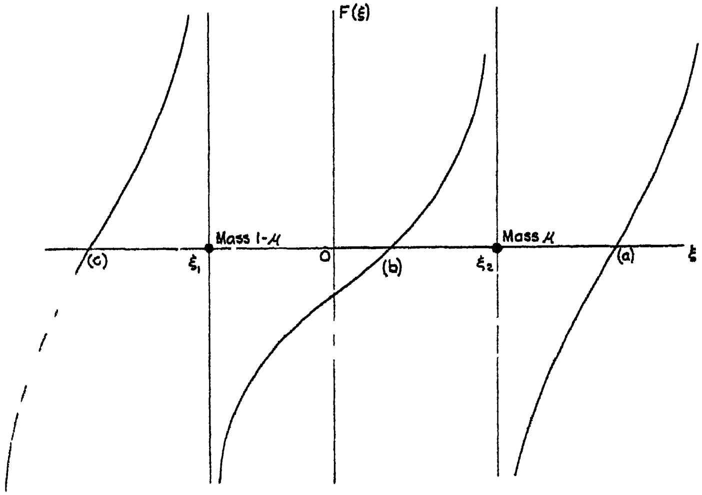


FIGURE 1

tive for $\xi = +\infty$, negative for $\xi = \xi_2 + \epsilon$, where ϵ is a small positive quantity, positive for $\xi = \xi_2 - \epsilon$, negative for $\xi = \xi_1 + \epsilon$, positive for $\xi = \xi_1 - \epsilon$ and negative for $\xi = -\infty$. Also $dF/d\xi$ is positive except at $\xi = \xi_1$ and $\xi = \xi_2$ where it becomes infinite. The graph of $F(\xi)$ is Figure 1.

Thus $F(\xi)$ changes sign three times by passing through zero at the following points

Point (a)	between	$+\infty$	and	ξ_2,
Point (b)	between	ξ_2	and	ξ_1,
Point (c)	between	ξ_1	and	$-\infty$.

Hence there are three straight-line equilibrium points if f, the force, varies as $1/r^n$, $n > 0$, as well as for f varying as $1/r^2$.

Point (a). Let the distance from the mass μ to the libration point on the ξ-axis between μ and $+\infty$ be denoted by $r_2^{(0)}$. Then

$$\xi - \xi_2 = r_2^{(0)}, \quad \xi - \xi_1 = 1 + r_2^{(0)}, \quad \xi = 1 - \mu + r_2^{(0)}.$$

The distance $r_2^{(0)}$ must satisfy

$$1 - \mu + r_2^{(0)} - \frac{1-\mu}{\left(1 + r_2^{(0)}\right)^n} - \frac{\mu}{r_2^{(0)n}} = 0,$$

or, when cleared of fractions,

$$\left[\left(1 - \mu + r_2^{(0)}\right)r_2^{(0)n} - \mu\right]\left(1 + r_2^{(0)}\right)^n - \left(1 - \mu\right)r_2^{(0)n} = 0. \quad (3)$$

The terms of lowest degree are

$$(n+1)r_2^{(0)n+1} - \mu - n\mu r_2^{(0)} = 0,$$

and by the theory of implicit functions equation (3) can be solved for $r_2^{(0)}$ as a power series in $\mu^{1/(n+1)}$ of the form

$$r_2^{(0)} = c_1\mu^{\frac{1}{n+1}} + c_2\mu^{\frac{2}{n+1}} + \ldots,$$

where

$$c_1 = \left(\frac{1}{n+1}\right)^{1/(n+1)}, \quad c_2 = \frac{n}{(n+1)^{(n+3)/(n+1)}},$$

converging for μ sufficiently small. This solution is real and unique (1) for n integral if the real root of $\mu^{1/(n+1)}$ is taken and (2) for $n = l/k$

provided the real root of $\mu^{k\ (l+k)}$ is taken for $l + k$ odd or the real positive root if $l + k$ is even.

Point (b). Let the distance from μ to the libration point on the ξ-axis between μ and $1 - \mu$ be denoted by $r_2^{(0)}$. Then

$$\xi - \xi_2 = -r_2^{(0)}, \quad \xi - \xi_1 = 1 - r_2^{(0)}, \quad \xi = 1 - \mu - r_2^{(0)}.$$

In this case the equation to be satisfied by $r_2^{(0)}$ is

$$1 - \mu - r_2^{(0)} - \frac{1-\mu}{(1 - r_2^{(0)})^n} + \frac{\mu}{r_2^{(0)n}} = 0,$$

or, when cleared of fractions,

$$[(1 - \mu - r_2^{(0)})r_2^{(0)n} + \mu](1 - r_2^{(0)})^n - (1 - \mu)r_2^{(0)n} = 0.$$

The terms of lowest degree in $r_2^{(0)}$ and μ are

$$-(1 + n)r_2^{(0)n+1} + \mu - nr_2^{(0)}\mu + \dots = 0,$$

and the solution for $r_2^{(0)}$ is

$$r_2^{(0)} = c_1\mu^{n+\bar{1}} - c_2\mu^{\overline{n+1}} + \dots,$$

where c_1 and c_2 have the same values as for point (a).

This solution will be unique if the same arrangements are made as to taking the $(n + 1)^{th}$ roots as at point (a).

Point (c). Let the distance from $1 - \mu$ to the libration point on the ξ-axis between the mass $1 - \mu$ and $-\infty$ be denoted by $r_1^{(0)}$.

The equation for $r_1^{(0)}$ vanishes for $\mu = 0$ and $r_1^{(0)} = 1$. We therefore put $r_1^{(0)} = 1 - \rho$. The value of ρ for the libration point (c) is found from

$$-(1 + \mu) + \rho + \frac{1-\mu}{(1-\rho)^n} + \frac{\mu}{(2-\rho)^n} = 0,$$

or, since $|\rho| \angle 1$,

$$(1 - 2^{n+1})\mu + 2^n(1 + n)\rho - n(2^n - \tfrac{1}{2})\rho\mu + \dots = 0.$$

Thus

$$\rho = \frac{2^{n+1} - 1}{2^n(1 + n)}\mu + \frac{n}{2^{2n+1}}\left(\frac{2^{n+1} - 1}{1 + n}\right)^2\mu^2 + \dots$$

a power series in μ converging for μ sufficiently small.

§ 1. *The Equilateral-Triangle Points*

To find the points not on the ξ-axis we must solve

$$\left.\begin{aligned} \xi - (1-\mu)\frac{\xi-\xi_1}{r_1^{n+1}} - \mu\frac{\xi-\xi_2}{r_2^{n+1}} &= 0, \\ \eta - (1-\mu)\frac{\eta}{r_1^{n+1}} - \mu\frac{\eta}{r_2^{n+1}} &= 0, \\ \zeta &= 0, \end{aligned}\right\} \qquad (4)$$

where $\eta \neq 0$. As in the case for $n = 2$, on dividing by η in the second equation of (4), we obtain

$$1 - \frac{1-\mu}{r_1^{n+1}} - \frac{\mu}{r_2^{n+1}} = 0.$$

Then multiplying this equation by $\xi - \xi_2$ and then by $\xi - \xi_1$ and subtracting the products separately from the first equation of (4) we arrive at the equations

$$\left.\begin{aligned} \xi_2 - (1-\mu)\frac{\xi_2-\xi_1}{r_1^{n+1}} &= 0, \\ \xi_1 - \frac{\mu(\xi_1-\xi_2)}{r_2^{n+1}} &= 0, \\ \zeta &= 0. \end{aligned}\right\} \qquad (5)$$

Now $\xi_2 = 1-\mu, \xi_1 = -\mu, \xi_2 - \xi_1 = 1$, and hence equations (5) reduce to

$$\left.\begin{aligned} 1 - \frac{1}{r_1^{n+1}} &= 0, \\ -1 + \frac{1}{r_2^{n+1}} &= 0, \\ \zeta &= 0. \end{aligned}\right\}$$

The only real solutions of these equations are $r_1 = 1$, $r_2 = 1$, $\zeta = 0$, and these points are the vertices of the equilateral triangles described on the line joining the centres of the finite masses and lying in the plane of revolution. Thus these vertices are points of libration for the more general force function as well as for the Newtonian law of attraction.

PART II

OSCILLATING SATELLITES NEAR THE STRAIGHT-LINE EQUILIBRIUM POINTS

§ 5. *The Differential Equations for Oscillations*

As in Moulton's *Periodic Orbits* we let

$$\left.\begin{array}{l} \xi = \xi_0 + \epsilon x, \ \eta = 0 + \epsilon y, \ \zeta = 0 + \epsilon z, \\ t - t_0 = (1 + \delta)\tau, \end{array}\right\} \qquad (6)$$

where ξ_0 denotes the straight-line equilibrium point (a), (b), or (c), ϵ is an arbitrary parameter, and δ a parameter to be determined so that x, y, z shall be periodic. The new independent variable is τ.

When we substitute (6) in (1), divide out by ϵ, and denote derivation with respect to τ by D we obtain

$$\left.\begin{array}{rl} D^2x - 2(1 + \delta)Dy = & (1 + \delta)^2 \, [X_1 + X_2\epsilon + X_3\epsilon^2 + \dots], \\ D^2y + 2(1 + \delta)Dx = & (1 + \delta)^2 \, [Y_1 + Y_2\epsilon + Y_3\epsilon^2 + \dots], \\ D^2z = & (1 + \delta)^2 \, [Z_1 + Z_2\epsilon + Z_3\epsilon^2 + \dots], \end{array}\right| \qquad (7)$$

$$X_1 = (1 + nA)x, \qquad X_2 = \frac{n+1}{2} B(-nx^2 + y^2 + z^2),$$

$$Y_1 = (1 - A)y, \qquad Y_2 = (n + 1)Bxy,$$

$$Z_1 = \quad -Az, \qquad Z_2 = (n + 1)Bxz,$$

$$X_3 = \frac{C}{6}\Big[n(n + 1)(n + 2)x^3 - 3(n + 1)(n + 2)\,x\,(y^2 + z^2) \Big],$$

$$Y_3 = \frac{n+1}{2} C[-(n + 2)x^2y + y^3 + yz^2],$$

$$Z_3 = \frac{n+1}{2} C[-(n + 2)x^2z + y^2z + z^3],$$

$$A = \frac{1-\mu}{[(\xi_0 + \mu)^2]^{\frac{n+1}{2}}} + \frac{\mu}{[(\xi_0 - 1 + \mu)^2]^{\frac{n+1}{2}}},$$

$$B = \begin{matrix}+\\+\\-\end{matrix} \frac{1-\mu}{[(\xi_0 + \mu)^2]^{\frac{n+2}{2}}} \begin{matrix}+\\-\\-\end{matrix} \frac{\mu}{[(\xi_0 - 1 + \mu)^2]^{\frac{n+2}{2}}},$$

$$C = \frac{1-\mu}{[(\xi_0 + \mu)^2]^{\frac{n+3}{2}}} + \frac{\mu}{[(\xi_0 - 1 + \mu)^2]^{\frac{n+3}{2}}}.$$

In B the upper, middle, or lower signs are to be used according as solutions in the vicinity of (a), (b), or (c) respectively are being treated.

Now let

$$x = \sum_{i=0}^{\infty} x_i \epsilon^i, \; y = \sum_{i=0}^{\infty} y_i \epsilon^i, \; z = \sum_{i=0}^{\infty} z_i \epsilon^i, \; \delta = \sum_{i=1}^{\infty} \delta_i \epsilon^i \tag{8}$$

be substituted in (7). The resulting equations will be cited as (7′). By equating the coefficients of the same powers of ϵ in (7′) we obtain sets of differential equations defining the various x_i, y_i, z_i. It will be shown that the arbitrary constants arising at each step and the δ_i can be chosen so that the x_i, y_i, z_i shall be periodic and shall satisfy certain assigned initial conditions.

§ 6. *The Generating Solutions*

By equating the coefficients of the terms independent of ϵ in (7′) we obtain

$$\begin{aligned} [D^2-(1+nA)]\,x_0 - 2Dy_0 &= 0, \\ 2Dx_0 + [D^2 - 1 + A]\,y_0 &= 0, \\ [D^2 - A\,]\,z_0 &= 0. \end{aligned} \tag{9}$$

The third equation is independent of the first two equations and has the solution

$$z_0 = C_1^{(0)} \cos\sqrt{A}\,\tau + C_2^{(0)} \sin\sqrt{A}\,\tau, \tag{10}$$

where $C_1^{(0)}$ and $C_2^{(0)}$ are arbitrary constants.

For the first two equations the determinant

$$\Delta = D^4 + D^2\,[2 - A(n-1)\,] + (1+nA)(1-A) \tag{11}$$

must vanish and therefore

$$D^2 = \tfrac{1}{2}\left[A(n-1) - 2 \pm \sqrt{[2 - A(n-1)\,]^2 - 4(1+nA)(1-A)}\,\right].$$

It may be shown as in section (82) of Moulton's (1920) Oscillating Satellites that

$$1 - A = (1-\mu)\left(1 - \frac{1}{{r_1^{(0)}}^{n+1}}\right)\left(1 - \frac{r_1^{(0)}}{r_2^{(0)}}\right) \angle\ 0, \text{ for point } (a),$$

$$1 - A = (1-\mu)\left(1 - \frac{1}{{r_1^{(0)}}^{n+1}}\right)\left(1 + \frac{r_1^{(0)}}{r_2^{(0)}}\right) \angle\ 0, \text{ for point } (b),$$

$$1 - A = \mu\left(1 - \frac{1}{{r_2^{(0)}}^{n+1}}\right)\left(1 - \frac{r_2^{(0)}}{r_1^{(0)}}\right) \angle\ 0, \text{ for point } (c).$$

If $n = l/k$, the real k^{th} root of ${r_1^{(0)}}^{(l+h)/k}$ is to be taken if k is odd and the real positive root if k is even. Hence the discriminant for D^2 is positive for all three equilibrium points and therefore D has two purely imaginary roots, $\pm\,\sigma i$ and two real roots $\pm\,\rho$.

The solutions of the first two equations of (9) are then

$$\left.\begin{aligned} x_0 &= K_1 e^{\sigma i\tau} + K_2 e^{-\sigma i\tau} + K_3 e^{\rho\tau} + K_4 e^{-\rho\tau}, \\ y_0 &= iM(K_1 e^{\sigma i\tau} - K_2 e^{-\sigma i\tau}) + N\ (K_3 e^{\rho\tau} - K_4 e^{-\rho\tau}), \end{aligned}\right\} \qquad (12)$$

where $K_1, \ldots, K_4$ are arbitrary, and

$$M = \frac{\sigma^2 + 1 + nA}{2\sigma} = \frac{2\sigma}{\sigma^2 + 1 - A},$$

$$N = \frac{\rho^2 - 1 - nA}{2\rho} = -\frac{2\rho}{\rho^2 - 1 + A}.$$

Equations (10) and (12) are generating solutions and yield three types of orbits, designated, as in the Oscillating Satellites (Moulton, 1920), as orbits of Class A, Class B, Class C.

§ 7. *Periodic Orbits of Class A*

The generating solutions for the orbits of Class A are

$$x_0 = 0, \ y_0 = 0, \ z_0 = \frac{1}{\sqrt{A}} \sin \sqrt{A}\tau.$$

The period is $2\pi/\sqrt{A}$. The desired solutions satisfying the initial conditions $z(0) = 0$, $\frac{dz(0)}{d\tau} = 1$, are

$$x_0 = y_0 = 0, \ z_0 = \frac{1}{\sqrt{A}} \sin \sqrt{A}\tau,$$

$$x_1 = \frac{(n+1)B}{4A}\left[-\frac{1}{1+nA} + \frac{(1+3A)\cos 2\sqrt{A}\tau}{1-(9-n)A+(12+3n)A^2}\right],$$

$$y_1 = -\frac{(n+1)\,B \sin 2\sqrt{A}\tau}{\sqrt{A}\,[1-(9-n)A+(12+3n)A^2]},$$

$$z_1 = \delta_1 = x_2 = y_2 = 0,$$

$$\delta_2 = \frac{-3(n+1)^2B^2[1-(5-n)A+(8+3n)A^2]}{16A^2(1+nA)[1-(9-n)A+(12+3n)A^2]} + \frac{3(n+1)C}{16A^2},$$

$$z_2 = \frac{n+1}{64A^{5/2}}\left[\frac{(n+1)(1+3A)B^2}{1-(9-n)A+(12+3n)A^2} - C\right]\left[3\sin\sqrt{A}\tau - \sin 3\sqrt{A}\tau\right],$$

$$x_{2j} = y_{2j} = z_{2j+1} = \delta_{2j+1} = 0,$$

$$x_{2j+1} = \sum_{k=0}^{j+1} C_k^{(2j+1)} \cos 2k\sqrt{A}\tau,$$

$$y_{2j+1} = \sum_{k=0}^{j+1} S_k^{(2j+1)} \sin 2k\sqrt{A}\tau,$$

$$z_{2j} = \sum_{k=0}^{j} S_k^{(2j)} \sin (2k+1)\sqrt{A}\tau,$$

where the C's and S's are known constants and

$$\sum_{k=0}^{j} (2k+1) S_k^{(2j)} = 0.$$

§ 8. *Periodic Orbits of Class B*

The generating solutions for orbits of Class B are, in trigonometric form,

$$\left.\begin{aligned} x_0 &= a_1^{(0)} \cos \sigma\tau + b_1^{(0)} \sin \sigma\tau, \\ y_0 &= M(b_1^{(0)} \cos \sigma\tau - a_1^{(0)} \sin \sigma\tau), \\ z_0 &= 0, \end{aligned}\right| \qquad (13)$$

$$M = \frac{\sigma^2 + 1 + nA}{2\sigma} = \frac{2\sigma}{\sigma^2 + 1 - A},$$

$a_1^{(0)}$, $b_1^{(0)}$ arbitrary constants.

The period of these solutions is $2\pi/\sigma$.

Since $z_0 = 0$ and since z is a factor of the third equation of (7), all the z_j will vanish. Thus the orbits of Class B are of two dimensions and lie wholly in the plane of revolution of the finite bodies.

The initial conditions

$$x(0) = 1, \ y(0) = 0 \qquad (14)$$

will be chosen for Class B orbits. There is no loss of generality in taking $x(0) = 1$ instead of an arbitrary constant as x carries the arbitrary constant ϵ in (6). Nor is there loss of generality in taking $y(0) = 0$ instead of an arbitrary constant as the infinitesimal must cross the η-axis for periodic motion and we may select the initial time $\tau = 0$ as the instant when the infinitesimal crosses the η-axis. When conditions (14) are imposed on x_j and y_j in (8), we obtain

$$\left.\begin{aligned} &x_0(0) = 1, \ x_j(0) = 0, \ j = 1, 2, 3, \ldots \\ &y_j(0) = 0, \ j = 0, 1, 2, 3, \ldots \end{aligned}\right\} \qquad (15)$$

As a result of (15) the general solutions (13) reduce to

$$x_0 = \cos \sigma\tau, \ y_0 = -M \sin \sigma\tau.$$

The orbits of Class B may be obtained by transforming the differential equations (7) to the normal form as in Buck's treatment (1920). As this is a somewhat laborious procedure we shall therefore construct the solutions directly.

Coefficients of ϵ *in* (12′). The differential equations at the second step are

$$\left.\begin{aligned} (D^2-1-nA)x_1-2Dy_1 &= X^{(1)}, \\ 2Dx_1+(D^2-1+A)y_1 &= Y^{(1)}, \end{aligned}\right\} \qquad (16)$$

$$X^{(1)} = 2\delta_1[Dy_0+(1+nA)x_0] - \frac{n+1}{2}B(nx_0^2-y_0^2),$$

$$Y^{(1)} = -2\delta_1[Dx_0-(1-A)y_0] + (n+1)Bx_0y_0.$$

The particular integrals of (16) in symbolic form are

$$x_1 = \frac{1}{\Delta}\begin{vmatrix} X^{(1)}, & -2D \\ Y^{(1)}, & D^2-1+A \end{vmatrix}, \quad y_1 = \frac{1}{\Delta}\begin{vmatrix} D^2-1-nA, & X^{(1)} \\ 2D, & Y^{(1)} \end{vmatrix}, \qquad (17)$$

where Δ is defined in (11). Since Δ vanishes for $D^2 = -\sigma^2$ the terms in $\cos\sigma\tau$ and also in $\sin\sigma\tau$ in (17) must vanish, otherwise the solutions for x_1 and y_1 would contain non-periodic terms in $\tau\,{}^{\sin}_{\cos}\,\sigma\tau$. Now δ_1 is a factor of the terms $\cos\sigma\tau$ and $\sin\sigma\tau$ in (17) and therefore we must put $\delta_1 = 0$. The complete solutions of (17) that satisfy the periodicity and initial conditions are

$$\begin{aligned} x_1 &= a_0^{(1)} + a_1^{(1)}\cos\sigma\tau + a_2^{(1)}\cos 2\sigma\tau, \\ y_1 &= -Ma_1^{(1)}\sin\sigma\tau + b_2^{(1)}\sin 2\sigma\tau, \\ \delta_1 &= 0, \end{aligned}$$

where $a_0^{(1)}, \ldots, b_2^{(1)}$ are known constants, and

$$\sum_{j=0}^{2} a_j^{(1)} = 0.$$

It will be necessary to carry the integrations one more step to show how δ_2 can be determined.

Coefficient of ϵ^2 *in* (7′). The differential equations at this step are

$$\left.\begin{aligned} (D^2-1-nA)x_2-2Dy_2 &= X^{(2)}, \\ 2Dx_2+(D^2-1+A)y_2 &= Y^{(2)}, \end{aligned}\right\} \qquad (18)$$

$$X^{(2)} = [2\delta_2(-M\sigma+1+nA)+g_1^{(2)}]\cos\sigma\tau + g_0^{(2)} + g_2^{(2)}\cos 2\sigma\tau + g_3^{(2)}\cos 3\sigma\tau,$$
$$Y^{(2)} = [2\delta_2\{\sigma-(1-A)M\} + h_1^{(2)}]\sin\sigma\tau + h_2^{(2)}\sin 2\sigma\tau + h_3^{(2)}\sin 3\sigma\tau,$$

where $g_0^{(2)}, \ldots, h_3^{(2)}$ are known constants. The particular integrals, in symbolic form, are

$$x_2 = \frac{1}{\Delta}\begin{vmatrix} X^{(2)}, & -2D \\ Y^{(2)}, & D^2-1+A \end{vmatrix}, \quad y_2 = \frac{1}{\Delta}\begin{vmatrix} D^2-1-nA, & X^{(2)} \\ 2D & Y^{(2)} \end{vmatrix}. \qquad (19)$$

As at the previous step, the coefficients of $\cos\sigma\tau$ and $\sin\sigma\tau$ in the above symbolic forms for the particular integrals must vanish as $\Delta = 0$ for $D^2 = -\sigma^2$ and would give rise to terms in $\tau \genfrac{}{}{0pt}{}{\cos}{\sin} \sigma\tau$. These coefficients, when equated to zero, are

$$\left.\begin{aligned}(-\sigma^2 - 1 + A)[\quad]_1 + 2\sigma[\quad]_2 &= 0,\\ 2\sigma[\quad]_1 - (\sigma^2 + 1 + nA)[\quad]_2 &= 0,\end{aligned}\right\} \qquad (20)$$

$$\begin{aligned}[\quad]_1 &= 2\delta_2(-M\sigma + 1 + nA),\\ [\quad]_2 &= 2\delta_2\{\sigma - M(1 - A)\}.\end{aligned}$$

The functional determinant of $[\quad]_1$ and $[\quad]_2$ in equations (20) is Δ with $D^2 = -\sigma^2$ and consequently vanishes. Hence the two equations in (20) are not independent and one of them, say the second, may be suppressed. Thus we obtain

$$\delta_2[\{4 - 2(n - 1)A\}\sigma^2 - 4(1 - A)(1 + nA)] + \text{constant} = 0. \qquad (21)$$

As the coefficient of δ_2 does not vanish we may solve (21) for δ_2 giving $\delta_2 = d_2$, a constant. With δ_2 thus determined the complete solutions of (18) satisfying the periodicity and initial conditions are

$$\begin{aligned}x_2 &= a_0^{(2)} + a_1^{(2)} \cos\sigma\tau + a_2^{(2)} \cos 2\sigma\tau + a_3^{(2)} \cos 3\sigma\tau,\\ y_2 &= -Ma_1^{(2)} \sin\sigma\tau + b_2^{(2)} \sin 2\sigma\tau + b_3^{(2)} \sin 3\sigma\tau,\\ \delta_2 &= d_2, \text{ a constant},\end{aligned}$$

where $a_0^{(2)}, \dots, b_3^{(2)}$ are known constants, and

$$\sum_{j=0}^{3} a_j^{(2)} = 0.$$

The remaining steps of the integration are similar to the preceding step. The differential equations at the step j are

$$\begin{aligned}(D^2 - 1 - nA)x_j - 2Dy_j &= X^{(j)},\\ 2Dx_j + (D^2 - 1 + A)y_j &= Y^{(j)},\end{aligned}$$

$$\begin{aligned}X^{(j)} = {} & [2\delta_j(-M\sigma + 1 + nA) + g_1^{(j)}]\cos\sigma\tau + g_0^{(j)}\\ & + g_2^{(j)}\cos 2\sigma\tau + \dots + g_{j+1}^{(j)}\cos(j + 1)\sigma\tau,\\ Y^{(j)} = {} & [2\delta_j\{\sigma - (1 - A)M\} + h_1^{(j)}]\sin\sigma\tau\\ & + h_2^{(j)}\sin 2\sigma\tau + \dots + h_{j+1}^{(j)}\sin(j + 1)\sigma\tau.\end{aligned}$$

The symbolic forms for the particular integrals are the same as (19) with $X^{(2)}$, $Y^{(2)}$ replaced by $X^{(j)}$, $Y^{(j)}$ respectively. Two equations similar to (20) must be satisfied by δ_j if the solutions are to be periodic. Now the functional determinant of these two equations is Δ and vanishes as at the previous step. Hence only one independent linear equation in δ_j remains and the coefficient of δ_j in this equation is the same as that of δ_2 in (21). Thus we find $\delta_j = d_j$, a constant, and the desired solutions at the step j have the form

$$\begin{aligned} x_j &= a_0^{(j)} + a_1^{(j)} \cos \sigma\tau + \ldots + a_{j+1}^{(j)} \cos (j+1)\sigma\tau, \\ y_j &= -Ma_1^{(j)} \sin \sigma\tau + \ \ldots \ + b_{j+1}^{(j)} \sin (j+1)\sigma\tau, \\ \delta_j &= d_j, \quad \sum_{k=0}^{j+1} a_k^{(j)} = 0, \end{aligned}$$

where $a_0^{(j)}, \ldots, b_{j+1}^{(j)}$ are known constants.

On substituting for the various x_j, y_j, δ_j in (8) and the results in (6) we obtain the two-dimensional orbits of Class B near the straight-line equilibrium points (a), (b), (c).

§ 9. *Periodic Orbits of Class C*

The generating solutions for the orbits of Class C are

$$\begin{aligned} x_0 &= a_1^{(0)} \cos \sigma\tau + b_1^{(0)} \sin \sigma\tau, \\ y_0 &= M(b_1^{(0)} \cos \sigma\tau - a_1^{(0)} \sin \sigma\tau), \\ z_0 &= \sin \sqrt{A}\,\tau. \end{aligned}$$

The period is $T = m_1 2\pi/\sigma = m_2 2\pi/\sqrt{A}$ where m_1 and m_2 are relatively prime integers and σ and $\sqrt{A}$ are assumed to be commensurable.

On account of the complexity of the problem in constructing orbits of Class C which are distinct from those of Class A or Class B when the law of attraction is Newtonian, no attempt will be made to construct Class C orbits for the more general force function.

§ 10. *Convergence of the Solutions*

The convergence of the solutions for the orbits of Class A and of Class B may be established by an existence theorem in which use is made of Poincaré's (1892, 1899) extension to Cauchy's theorem. On the other hand, we may apply MacMillan's (1913) theorem where it is shown that if the formal construction of periodic solutions can be made then such solutions will converge for all values of the time provided the parameter is sufficiently small numerically.

PART III

OSCILLATING SATELLITES NEAR THE EQUILATERAL-TRIANGLE EQUILIBRIUM POINTS

§ 11. *The Differential Equations for Oscillations*

The co-ordinates of the equilateral triangle equilibrium points are

$$\begin{aligned} &\text{Point I,} \quad \xi_0 = \tfrac{1}{2} - \mu, \ \eta_0 = \tfrac{1}{2}\sqrt{3}, \ \zeta_0 = 0, \\ &\text{Point II,} \quad \xi_0 = \tfrac{1}{2} - \mu, \ \eta_0 = -\tfrac{1}{2}\sqrt{3}, \ \zeta_0 = 0. \end{aligned} \qquad (22)$$

Only the periodic oscillations near one of these points need be considered in detail, as the periodic oscillations near the other point may be obtained by changing the sign of $\sqrt{3}$ in the solutions near the point first considered. We shall select Point I for detailed consideration.

The case for $n = 2$ has already been considered by Buck As in his treatment of the problem we put

$$\xi = \tfrac{1}{2} - \mu + \epsilon x,\ \eta = \tfrac{1}{2}\sqrt{3} + \epsilon y,\ \zeta = \epsilon z \qquad (23)$$
$$t - t_0 = (1 + \delta)\tau.$$

On substituting (23) in (6) and dividing out ϵ we obtain the following differential equations defining the oscillations near Point I:

$$\left.\begin{aligned}
&\left[D^2 - \frac{n+1}{4}(1+\delta)^2\right]x - \left[2(1+\delta)D + \frac{(n+1)(1-2\mu)\sqrt{3}(1+\delta)^2}{4}\right]y \\
&\qquad = (1+\delta)^2[\epsilon X_2 + \epsilon^2 X_3 + \ldots],\\
&\left[2(1+\delta)D - \frac{n+1\,(1-2\mu)\sqrt{3}(1+\delta)^2}{4}\right]x + \\
&\qquad [D^2 - \tfrac{3}{4}\,(n+1)(1+\delta)^2]\ y \\
&\qquad = (1+\delta)^2\ [\epsilon Y_2 + \epsilon^2 Y_3 + \ldots\ ],\\
&D^2 + (1+\delta)^2]\,z = (1+\delta)^2\,[\,\epsilon Z_2 + \epsilon^2 Z_3 + \ldots],
\end{aligned}\right\} \qquad (24)$$

$$X_2 = \frac{-(1-2\mu)}{16}\left[(n+1)(n-9)x^2 + (n+1)(3n+5)y^2 - 4(n+1)z^2\right]$$
$$- \frac{(n+1)(n-1)}{8}\sqrt{3}xy,$$

$$Y_2 = \frac{-(n+1)(n-1)\sqrt{3}}{16}(x^2 + 3y^2) + \frac{(n+1)\sqrt{3}}{4}z^2$$
$$- \frac{(1-2\mu)(n+1)(3n+5)}{8}xy,$$

$$Z_2 = \frac{n+1}{2}\,z\,[(1-2\mu)x + \sqrt{3}y\,].$$

The remaining X_j, Y_j, Z_j are polynomials of degree j in x, y, and z. The Z_j carry the factor z.

§ 12. *The Equations of Variation and Their Solutions*

In order to integrate (24) we put

$$x = \sum_{i=0}^{\infty} x_i\epsilon^i,\ y = \sum_{i=0}^{\infty} y_i\epsilon^i,\ z = \sum_{i=0}^{\infty} z_i\epsilon^i,\ \delta = \sum_{i=1}^{\infty} \delta_i\epsilon^i. \qquad (25)$$

Let these substitutions be made in (24) and let the resulting equations be cited as (24′). The terms in (24′) that are independent of ϵ are

$$\left.\begin{aligned} D_1x_0+D_2y_0 &= 0, \\ D_3x_0+D_4y_0 &= 0, \\ (D^2+1)\,z &= 0, \end{aligned}\right| \qquad (26)$$

$$D_1 = D^2 - \tfrac{1}{4}\,(n+1),\ D_2 = -\,2D - \tfrac{1}{4}\,(n+1)(1-2\mu)\sqrt{3},$$
$$D_3 = 2D - \tfrac{1}{4}\,(n+1)(1-2\mu)\sqrt{3},\ D_4 = D^2 - \tfrac{3}{4}\,(n+1).$$

These equations are known as the *equations of variation.*

In order to find the solutions of (26, a, b), that is, the first two equations of (26), we must first obtain the roots of the determinant $D_1D_4 - D_2D_3 = 0$ or

$$D^4+(3-n)D^2+\tfrac{3}{4}\,(n+1)^2\mu(1-\mu) = 0. \qquad (27)$$

Suppose the roots of this equation are $\pm\,\rho_1$, $\pm\,\rho_2$. Then the solutions of (26, a, b) are

$$x_1 = a_1e^{\rho_1\tau}+a_2e^{-\rho_1\tau}+a_3e^{\rho_2\tau}+a_4e^{-\rho_2\tau},$$
$$y_1 = b_1\,a_1e^{\rho_1\tau}+b_2a_2e^{-\rho_1\tau}+b_3a_3e^{\rho_2\tau}+b_4a_4e^{-\rho_2\tau},$$

where the a's are arbitrary and the b's are known constants.

The solution of (26, c) is

$$z_0 = c_1\sin\tau + c_2\cos\tau.$$

§ 13. *The Orbits in the Plane*

The orbits in the xy-plane will exist only when either ρ_1 or ρ_2 is purely imaginary. Now

$$\rho_1^2 \text{ or } \rho_2^2 = \frac{n-3\pm\sqrt{(3-n)^2-3(n+1)^2\mu(1-\mu)}}{2},$$

and ρ_1 or ρ_2 will be purely imaginary if the expression for ρ_1^2 or ρ_2^2 is negative and real. In order that ρ_1^2 or ρ_2^2 shall be negative we must have $n<3$ and to be real

$$(3-n)^2 \geqq 3(n+1)^2\mu(1-\mu),$$

or $\mu(1-\mu)$ must not exceed $\dfrac{(3-n)^2}{3(n+1)^2}$ or μ must not exceed

$$1-\sqrt{1-\frac{4(3-n)^2}{3(1+n)^2}}\,.$$

This limitation upon μ is real only for $n > 15-8\sqrt{3}$ or $n > 1.144$. For example if $n = 1$ there is no limitation upon μ other than, by notation, it must not exceed 1/2. If $n = 2$ then μ must be less than .0385 . . . as first found by Lagrange in 1772.

The plane orbits for $n = 2$ were obtained by Buck. The construction of orbits for $0 < n < 3$ is essentially the same as for $n = 2$ and the work will not be given here.

§ 14. *The Three-Dimensional Orbits*

It will be assumed that neither ρ_1 nor ρ_2 is commensurable with $\sqrt{-1}$ for n in the range $0 < n < 3$. This assumption need not be made for $n > 3$ as then the ρ's are no longer purely imaginary.

The generating solutions for the three-dimensional orbits are

$$x_0 = y_0 = 0,\ z_0 = a_1 \sin \tau + a_2 \cos \tau,$$

and the period is 2π.

Initial conditions $z(0) = 0$, $Dz(0) = 1$ will be selected. Then

$$\left.\begin{array}{l} z_j(0) = 0,\ j = 0, 1, 2, \ldots, \\ Dz_0(0) = 1,\ Dz_j(0) = 0,\ j = 1, 2, 3, \ldots \end{array}\right\} \qquad (28)$$

Step 1. The solutions at the first step satisfying (28) are

$$x_0 = y_0 = 0,\ z_0 = \sin \tau.$$

Step 2. Coefficients of ϵ *in* (24′).

The differential equation in z_1 will be considered first. It is

$$(D^2 + 1)z_1 = -2\delta_1 \sin \tau.$$

The periodic solution satisfying (28) is

$$z_1 = 0,\ \delta_1 = 0.$$

The differential equations in x_1 and y_1 are

$$D_1 x_1 + D_2 y_1 = X^{(1)} = \frac{(1 - 2\mu)(n + 1)}{8}(1 - \cos 2\tau),$$

$$D_3 x_1 + D_4 y_1 = Y^{(1)} = \tfrac{1}{8}(n + 1)\sqrt{3}\,(1 - \cos 2\tau),$$

and the desired solutions are

$$x_1 = a_0^{(1)} + a_2^{(1)} \cos 2\tau,$$
$$y_1 = b_0^{(1)} + b_2^{(1)} \cos 2\tau,$$

where $a_0^{(1)}, \ldots, b_2^{(1)}$ are known constants.

Step 3. One more step of the integrations will show how the process may be carried on to any desired length.

We shall consider the z_2- equation first. It is

$$(D^2 + 1)z_2 = -2\delta_2 z_0 + \frac{n+1}{2} z_0[(1 - 2\mu)x_1 + \sqrt{3}y_1]$$

$$= \sin \tau\left[-2\delta_2 + \frac{n+1}{2}\left\{a_0^{(1)}(1 - 2\mu) + \sqrt{3}\,b_0^{(1)} - \frac{(1 - 2\mu)}{2}a_2^{(1)} + \frac{\sqrt{3}}{2}b_2^{(1)}\right\}\right] \qquad (29)$$

$$+ \sin 3\tau\left[\frac{n+1}{4}\left\{(1 - 2\mu)a_2^{(1)} + \sqrt{3}\,b_2^{(1)}\right.\right].$$

In order that z_2 may be periodic, the coefficient of sin τ in (29) must vanish. This determines $\delta_2 = d_2$, say, and the desired solution for z_2 is

$$z_2 = C_3^{(2)}(-\sin\tau + \sin 3\tau),$$

where $C_3^{(2)}$ is a known constant.

The equations in x_2 and y_2 are

$$\begin{aligned} D_1 x_2 + D_2 y_2 &= 0, \\ D_3 x_2 + D_4 y_2 &= 0, \end{aligned}$$

and the desired solutions are

$$x_2 = y_2 = 0.$$

At the general step j, if $j = 2k + 1$, the solutions are

$$\begin{aligned} x_{2k+1} &= a_0^{(2k+1)} + a_2^{(2k+1)}\cos 2\tau + \ldots + a_{2k+2}^{(2k+1)} \cos (2k+2)\tau, \\ y_{2k+1} &= b_0^{(2k+1)} + b_2^{(2k+1)}\cos 2\tau + \ldots + b_{2k+2}^{(2k+1)} \cos (2k+2)\tau, \\ z_{2k+1} &= \delta_{2k+1} = 0; \end{aligned}$$

and if $j = 2k$ the solutions are

$$\begin{aligned} x_{2k} &= y_{2k} = 0, \\ z_{2k} &= C_1^{(2k)} \sin\tau + C_3^{(2k)} \sin 3\tau + \ldots + C_{2k+1}^{(2k)} \sin (2k+1)\tau, \\ \delta_{2k} &= d_{2k}, \text{ constant}, \end{aligned}$$

$$C_1^{(2k)} + 3C_3^{(2k)} + \ldots + (2k+1)\, C_{2k+1}^{(2k)} = 0,$$

where the a's, b's, C's and d's are known constants.

By substituting for the various x_j, y_j, z_j, δ_j in (25) and the results in (23) we find the equations for the three dimensional orbits. The infinitesimal is initially projected from the origin along the ζ-axis with a velocity that is proportional to ϵ.

§ 15. *The Convergence of the Solutions*

As in Part II, section 9, the solutions may be shown to be convergent by applying MacMillan's theorem (1913).

University of British Columbia

REFERENCES

Buck, T. Chapter ix in Moulton's *Periodic Orbits.* Carnegie Institution, Washington, D.C. 1920.

Lagrange, J. L. Collected Works.

MacMillan, W. D. Trans. Amer. Math. Soc., 13: 146-58 (1913).

Moulton, F. R. *Celestial Mechanics.* Chapter viii. Washington. 1914.

——— *Periodic Orbits.* Chapter v. Carnegie Institution, Washington, D.C. 1920.

Poincaré, H. *Les méthodes nouvelles de la mécanique céleste.* Paris. 1892, 1899.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLI : SERIES III : MAY, 1947
SECTION THREE

Illumination for Concealment of Ships at Night[1]

By E. GODFREY BURR

Presented by A. NORMAN SHAW, F.R.S.C.

§ 1. *Experiments preceding the Concealment Scheme*

During the fall of 1940 the writer and H. Mortimer[2] were studying the visibility range of aircraft at night, particularly with a view to the extension of that range by the aid of large aperture sighting telescopes of considerable magnification.

Binoculars of this type of the highest quality give a magnified image with a reduction of field brightness of about 50 per cent and with but a small reduction of contrast between the dark image or silhouette and the field. Under these circumstances the range of visibility at which the presence of the image can be noted is greater than with the naked eye in a ratio of the order of 70 per cent of the magnification ratio. Any considerable scattered light within the optical system directly reduces the contrast of the silhouette and very materially reduces this range ratio.

During this work the writer noted the extreme contrast obtained in the telescope field when an aircraft was viewed flying over an airport in open country. The silhouette had a velvety black appearance. He became impressed with the idea that the increase of visibility range obtained by the night glass could be defeated by an artificial reduction of contrast obtained by illumination of the surfaces viewed in the silhouette range. Since a diffusely emitting surface is equally bright in all directions a reduction of contrast to an approximate brightness-match with the background would be equally effective for all angles between the surface and the observer. The power requirement for obtaining this match with incandescent lamps was found to be quite reasonable for all night conditions.

A striking and conclusive demonstration occurred on December 4, 1940. The airport was covered with fresh snow and the moon was rising. Lightly wooded country surrounded the airport. A plane approaching at low altitude in level flight over the woods was clearly sighted with the night glass at extreme range and followed in. It suddenly disappeared at the edge of the airport and was not seen again until right overhead when with the naked eye it showed brighter than the sky background. The reflected moonlight from the snow had produced an approximate brightness-match of the plane with the sky.

---

[1] Permission to publish granted by the Director of Naval Intelligence, Naval Service Headquarters, Ottawa.

[2] Hector Mortimer, F.R.C.S. (Ed.), Research Fellow in Endocrinology, McGill University.

The technical difficulties of illuminating aircraft surfaces unless translucent skins could be developed led to the consideration of application to ships.

§ 2. *Demonstration of Concealment Effect on Ships*

A few days later at a meeting at Royal Canadian Navy Headquarters arranged for the writer by the Acting President of the National Research Council of Canada, it was disclosed that the usual method of submarine attack on our convoys was at night and that the submarine approached on the surface using night glasses. Such attacks were delivered generally from beyond the range at which our "lookouts" could sight the submarine. It was felt that any considerable reduction of visibility of the target ships would be a valuable defence measure. After laboratory demonstrations, a full-scale demonstration was arranged using, with simple rheostatic control, green-coloured incandescent lamp bulbs in commercial sign-lighting fixtures on temporary supports installed on the corvette H.M.C.S. *Cobalt*. This sea demonstration, held on January 22, 1941, outside Halifax, resulted in instructions to attempt the development of a reasonably seaworthy installation. The demonstration equipment was then shipped to the Admiralty and, after a trial, they proceeded with further development.

§ 3. *First Phase of Development*

During the next three months lighting units giving a fan-shaped beam by use of commercial prismatic plates and incorporating a blue-green glass filter to cut off the excess red, were made and installed on the corvette H.M.C.S. *Chambly*, using retractable pipe-frame supports. With this installation, which was used for frequent trials, some in quite rough weather, the general means of meeting the practical problems of shielding the sources from view under all angles of vision and with the ship pitching or rolling were defined. Also the mechanical problem of support for the hull units, which would project overside, was judged to be readily solvable without excessive weight of structures. Doubt of the mechanical feasibility continued to agitate naval officers with some exceptions until it was proved in 1943 by the *Edmunston* installation.

This stage of development was terminated before the end of May, 1941, when *Chambly* was ordered to sea, but it was by this time evident that range reduction by one-half could be expected under all conditions tested and under very favourable conditions the range might be reduced to one-quarter of the ordinary range.

§ 4. *Second Phase of Development: Defence of Merchant Ships*

The writer then obtained permission to disclose this invention to the U.S. Navy and at a meeting June 5, 1941, at the Bureau of Ships, Washington, they decided to develop equipment for demonstration on a large merchant ship which had been indicated by Admiralty as a desirable application. The General Electric Company were given an order to produce suitable equipment and the writer was advised to collaborate closely with the U.S. Navy in the project. A very compact lighting unit was developed using a reflector with cylindrical and parabolic curvatures in transverse and longitudinal vertical planes respectively, which gave a fan-shaped beam and nearly uniform illumination over a considerable area of a plane illuminated at a general angle of incidence of 15° when spaced about 15 feet apart. Equipped with a 40-watt 6-volt automobile type lamp, a blue-green filter of thick glass, and a housing of ordinary waterproof type, it formed a reasonably practical unit which was used in all later work. Installed as above, at top voltage the brightness of ship surfaces of reflectance 20 per cent attained the values of the night sky on moonlight nights.

In March, 1941, the Optics Department of the National Research Council of Canada had started to make an experimental photoelectric automatic control to test the feasibility of such a device. The extreme difficulty of maintaining insulation of the photoelectric cells to allow of balancing for sky and ship was quite evident and before it was ready for trial the General Electric Company, using our experience, constructed a control unit employing the newly developed multiplier photo tube and a rotating shutter exposing the photo cell alternatively to the ship surface and the sky. The alternating current thus produced when a brightness unbalance existed, was amplified and used to restore the balance through the means of Thyratron control of saturating reactors in the alternating current lamp circuit. The control was dual for port and starboard sides and was found to be sufficiently sensitive for practical development.

This control and the lighting units were installed by the General Electric Company on a large freighter in the U.S. Navy service and trials were run off out of Boston on New Year's Eve, 1942. The results were satisfactory in spite of the fact that the ship being empty of freight the undercurvature of the hull was exposed and no light could reach it. This was not a normal condition for loaded ships in convoy. All upper parts disappeared at quite close ranges but this dark shadow persisted and reduced the range reduction ratio to about one-half.

The United States, newly engaged in war, could not find justification for the large industrial output needed to develop this device for defence of merchant ships. They made the next spring a partial installation on an anti-submarine patrol vessel, evidently with the offensive aspect in view, but this was abortive because the overside hull units were ordered omitted, as at this time it appeared to them difficult to make these overside units seaworthy. The hull forms the largest and most compact part of the silhouette area, and so mainly determines the visibility range.

§ 5. *Third Phase of Development: Offensive Use in Anti-Submarine Warfare*

In midsummer, 1942, a definite step to make possible tests of the offensive value of the invention was taken. The Royal Canadian Navy authorized the writer to acquire a number of the General Electric Company's lighting units and to proceed with an installation for two corvettes at the time of their refit. The General Electric Company provided the complete photoelectric control applied to a 115-volt 60-cycle multiple distribution system with individual transformers down to 6 volts at each lighting unit. This installation is illustrated in the plates. The ships equipped were H.M.C.S. *Edmundston* and H.M.C.S. *Rimouski*. About sixty units were used and the maximum power consumption was 2.5 kilowatts. The blue-green light filters on the General Electric units were known to be too green, showing a residual green colour contrast against the sky in full moonlight, but the state of industry and the instructions given made it impossible to provide more suitable ones and this defect was apparent later under the conditions given above.

Delays in refit postponed the trials of *Edmundston* until June, 1943, and *Rimouski* was still not ready.

The *Edmundston* trials were very encouraging, range reduction by 50 to 70 per cent being reported by the Director of Anti-Submarine

Warfare, British Admiralty Mission, Washington, who witnessed the trials from a submarine. Earlier phases of the same trials were attended by a representative of the Department of Scientific Research, Admiralty, which department had carried out considerable independent development following communication of the invention to Admiralty in February, 1941. The writer was unable to participate in this work.

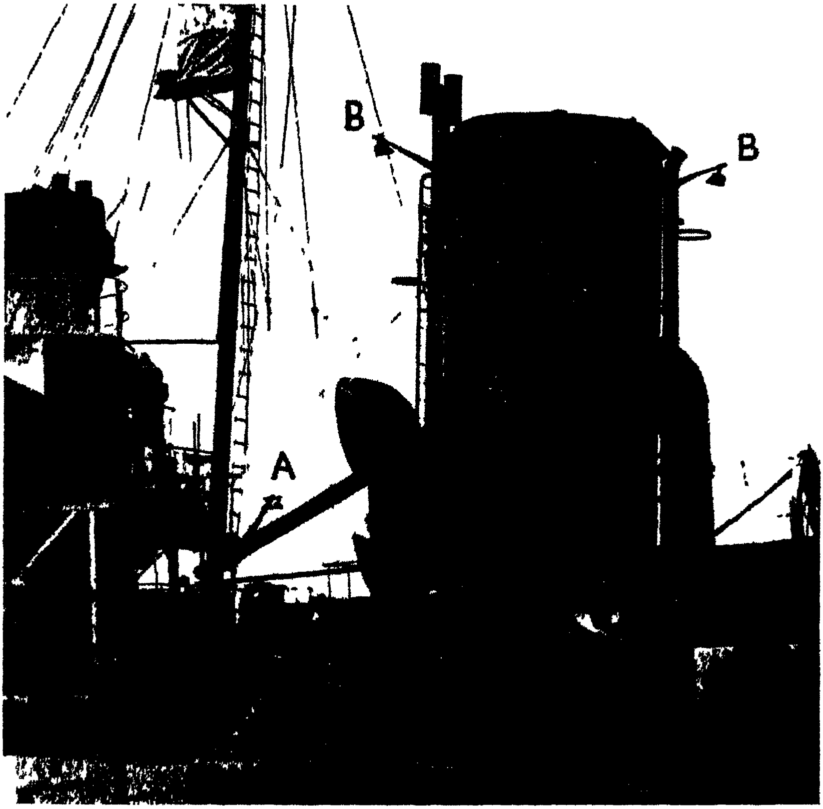


FIGURE I.—INSTALLATION ON CORVETTE

A. Overside midships unit in place.
B. Funnel unit in place.
C. Fixed superstructure unit.

Following these trials *Edmundston* proceeded on her normal tour of operational sea duty, much of it in the Bay of Biscay. A special officer accompanied her and operating experience was gained during the next six months, when he was called back to report Later, at the time of normal refit, the supporting brackets were found to have endured the stress of a normal tour of sea duty as well or better than the usual deck fittings of a warship.

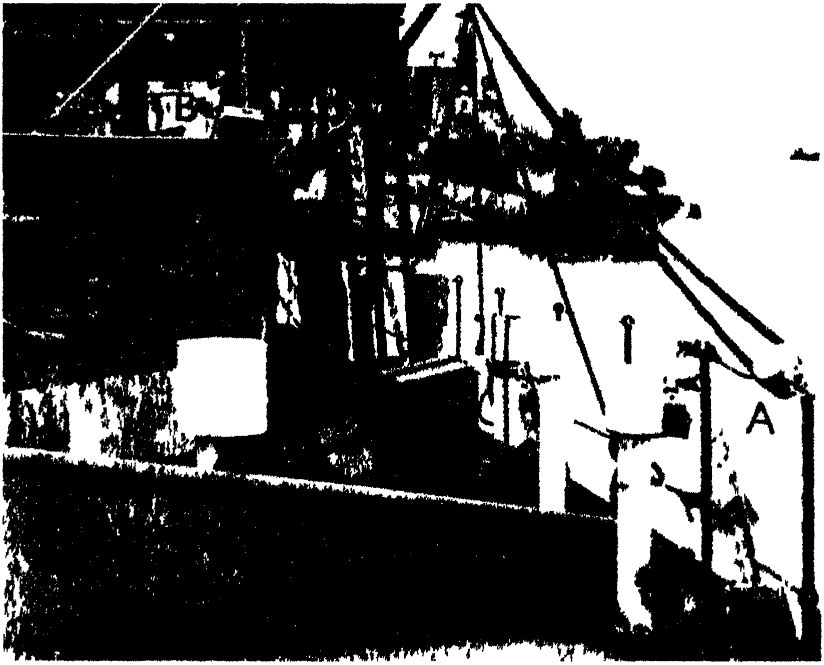


FIGURE II.—INSTALLATION ON CORVETTE.

A. Overside bow unit "stowed."
B. Overside midships unit "stowed."

Rimouski completed refit in October and operated with an escort group in Atlantic convoy using the device most of the time under the direction of a special officer, who thus gained operating experience and discovered the defects of the trial installation, until he was recalled three months later. At the end of her normal tour of duty the brackets were found in good condition, but as in the case of *Edmundston*, improvement of control gear and of the lighting units was needed to give a reliable installation, free from the necessity of frequent maintenance.

At an Admiralty meeting in London in August, 1943, to which the writer was called, proposed staff requirements were laid down. The ones which were not met by the existing installations mainly concerned the control gear. This should be automatic, with an alternative visual control. It should be trainable to give the balance most accurately for the actual direction of a known enemy. Also it was decided that the most suitable class of ship was the frigate.

§ 6. *Fourth Phase of Development: Improvements for Anti-Submarine Warfare*

This required a complete redesign of the whole control gear. Also, experience on the corvettes had shown that when the ship was heavily rolling, the "gimball" support, used to give horizontal line of sight, was unequal to the task. It was decided to incorporate gyroscopic stabilization. The Electric Engineering Department of the National Research Council undertook this task of redesign and construction of a prototype equipment.

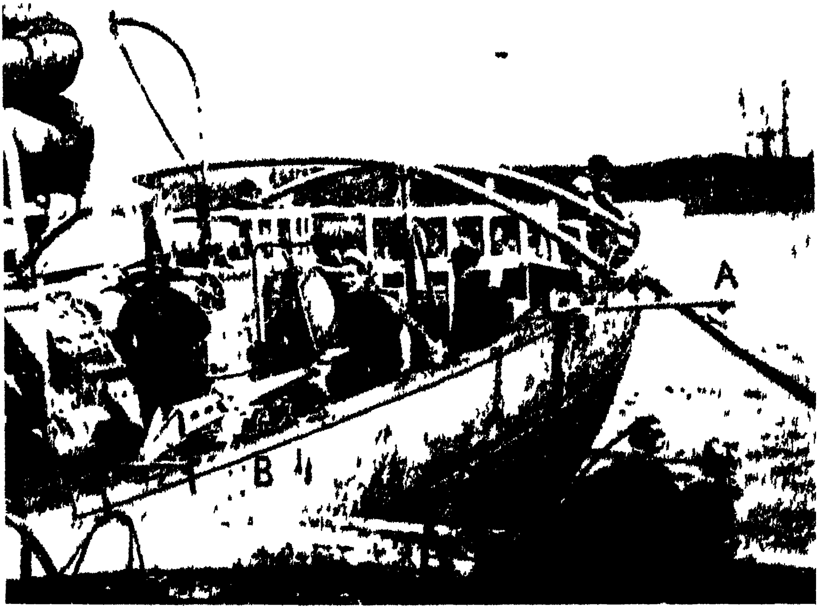


FIGURE III.—INSTALLATION ON CORVETTE.

A. Overside stern unit in place.
B. Overside stern unit partly "stowed."

A project for eight frigates was undertaken in Canada, six of which would form a Canadian escort group and two be attached to the Royal Navy.

The lighting units were redesigned to be waterproof against pressure and a blue filter replaced the 1941 design which gave the residual green colour contrast under certain conditions, as previously noted. The whole design was much improved mechanically, electrically, and optically based on the experience gained during the sea duty of the two corvettes.

To facilitate installation, a means was found for adjusting the units in a large shore laboratory. The whole of the surface of the hull of a frigate and parts of the superstructure were reproduced by plywood panels on a wood framework corresponding to the ribs of the hull. One side of the room had a fixed set-up of rather more than one-quarter of the ship's length aft from the bow. The remaining three-quarters of the ship could be reproduced on the other side of the laboratory one-quarter at a time, and changes required only a few hours.

The whole laboratory, a disused electrical sub-station machinery room, was adequately darkened and full-scale studies of the best spacing and positions of the lighting units could be made at any time. It was the intention to check and adjust the newly designed units in this laboratory and then label them for installation in shipboard.

This laboratory was used in the spring of 1944 to obtain in full scale the type of lighting distribution which was indicated as most desirable after several months of study by the author of the residual pattern of the frigate silhouette on a miniature scale of 1/16 inch equals 1 ft.[3] For this, a laboratory at McGill University, which had been equipped and calibrated by the author in 1942,[4] was available and was used with minor change.

The National Research Council had completed a prototype of a gyro-stabilized trainable automatic and visual control by the fall of 1944, and this was installed for the purpose of sea trials on *Rimouski* during her refit, replacing the earlier control used in her operational duty. The old lighting units were not replaced as the production of the new design had not started. Trials on this new gear were held out of Halifax in the late fall, 1944, at Bermuda in February, 1945, and in Scotland in April, 1945.

---

[3] E. G. Burr, "R.C.N. laboratory experiments on diffusion lighting of frigate." R.C.N., DTR 100, 1 April 1944.

[4] E. G. Burr, and A. S. Elliott, "Studies in night visual acuity." R.C.A.F., AM 25, 28 November 1942.

It appeared that the new control gear was much more suitable and should be adequate for the purpose.

By the end of 1944, the approaching end of the war had caused the abandonment of the project and the Bermuda and Scotland trials were its conclusion.

§ 7. *Notes on Sea Trials*

At the *Edmundston* trials (midsummer, 1943), which extended over several nights in a period of two weeks, some dramatic moments occurred.

With full moonlight *Edmundston* was viewed standing directly in the bright moon path, as a black object visible with the naked eye. With light switched on, she disappeared and could not be seen with 7 × 50 light glasses. These glasses in clear atmosphere have a range increase ratio of more than four.

On a very dark night with intermittent rain, which prevents efficient use of night glasses, the observing submarine having with the naked eye sighted the control ship (a similar corvette painted with the lately developed "Western Approaches" camouflage) at above 700 yards approached closer, but could not find *Edmundston.* Fearing too close approach, the submarine stopped and ordered "Lights-off." *Edmundston* was off the bow at 300 yards where a very careful watch had been kept. When "Lights-on" was again ordered, she was completely invisible, but after a very close scrutiny with night glasses she was just discernible.

The next day, an officer of *Edmundston* asked what the submarine had been doing all night coming up, stopping, and turning away. Asked if he had seen the submarine, he said he could see her "wake" clearly. Evidently on that occasion, the submarine was detected at ranges when *Edmunston* was invisible from the submarine.

The same officer had previously been told by the German captain of a captured submarine that they had one great advantage, the submarine could always see a corvette first.

On another occasion, in staged sea trials, three experienced officers in a motor launch could just see *Rimouski* at a range of 7,000 yards with night glasses when the lights were off. Signalling "Lights-on," they approached rapidly on the known bearing and first sighted at 1,200 yards. During the considerable time of the approach run of three sea miles, readings of sky brightness were taken and a slight darkening of the sky occurred which would reduce the ratio to five.

On another occasion, a frigate, unaware that the tests were being made, approached *Rimouski*, being first located by radar. She was sighted at 4,000 yards. She passed within 100 yards and was then challenged. The frigate captain the next day stated he had been challenged by *Rimouski*, but did not see her at any time and thought her to be a mile or two away.

McGill University.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLI : SERIES III : MAY, 1947
SECTION THREE

Transverse Waves in Plane Soap Films

By JOHN SATTERLY, F.R.S.C.

Synopsis

Transverse waves were propagated along narrow rectangular soap films, the source being an electromagnetically excited dipper penetrating the film near one end. Frequencies varying from 30 to 400 per second were used. Standing waves were perceived by eye using the reflection of an extended source of light at large angles of incidence to detect the presence of crests and troughs. To detect travelling waves a phonic wheel rotating with the same frequency as the dipper and provided with slots for vision was used. With this arrangement the pattern of the waves is seen at rest.

Travelling waves and stationary waves were observed with films of different lengths and breadths and their wave-lengths measured approximately. Noteworthy features were (1) the crests are sometimes inclined to the length of the film by an angle as large as 45°; (2) an X-shaped pattern was sometimes observed giving possible evidence of the simultaneous occurrence of a beam and its reflected beam from the far end of the film; (3) by using an L-shaped film and a reflecting edge of mica pushed partially into the film at the corner at an angle of 45°, the beam coming down one arm of the L could be reflected down the other arm.

These experiments may have important analogies in the propagation of electromagnetic waves along surfaces of metals and in the action of so-called "wave-guides."

The Experiments

A common lecture experiment is to place a plane soap film across the flanged end of a short length of a three-inch pipe and use it to reflect a bright beam of light from an arc-lantern. By means of a large lens a magnified image is produced on the screen. As the film drains interference occurs between the beams reflected from the front and back surfaces of the film and brilliantly coloured horizontal bands appear on the screen. Weaker soap solutions often give brighter colours but the film is not so lasting as with stronger solutions. If now a narrow stream of air be directed upon the film by means of a glass nozzle the horizontal bands of colour break up into gorgeously coloured vortices.

Also, if a high-pitched *note* be sung into the supporting tube the film is set in vibration, the pattern on the screen showing stationary sets of criss-crossing ripples. The higher the frequency of the note the finer the crispations of the pattern; the appearance is similar to that of a Chladni plate executing high frequency vibrations. The experiment with this unexpected burst of song and the changing pattern on the screen usually "brings down the house."

C. V. Boys in his little book entitled *Soap Bubbles* discusses the colours of these films and gives a diagram showing the relation between the colour, its position in Newton's scale of colours, and the thickness of the film. But he says little on vibrations. A. S. C. Lawrence in his book on *Soap Films* gives photographs of draining films showing the horizontal bands and also photographs of the vortices produced by an impinging jet of air, and makes the brief remark that startling changes are produced by very small fluctuations of air pressure in the jet.

Hartmann and Mathes[1] have described experiments on "Vibrations of a Liquid Membrane Exposed to High Frequency Sound Waves." The source of sound was an Acoustic Generator giving a note of frequency as high as 12,400 per second. The liquid membrane was a *small* soap film of diameter 7 mm. when circular in shape and of 4 × 1.2 mm. dimensions when rectangular. Its thickness was approximately 1 to 3 hundredths of a mm. The photographs reveal that the impinging high frequency waves set the film in stationary undulation showing at first the existence of ripples and finally, as the film drained, minute granulations. After considering the mechanics of the problem and the variations seen in the photographs they conclude (p. 891) that "the natural state of vibration of a liquid film is thus obviously that to which three sets of stationary transverse waves under 120° with each other would give rise" and that "if the film is too thick the amplitude of the transverse waves is too small to be observed. When the thickness of the film has become too small the film breaks. In between there is no change in the length of the stationary wave with thickness, though this is a point to be cleared up."

Attempts at calculations connecting wave-length and frequency were made using the simple equation for standing waves in stretched strings and the Kelvin-Rayleigh expression for surface capillary waves.

---

[1] *Phil. Mag.*, **22**: 883 (1936).

As with Hartmann we may attempt to treat a long narrow film as a stretched string and apply the equation for travelling or standing waves namely

$$v = \sqrt{\frac{\text{stretching force}}{\text{mass per unit length}}}$$

where v = velocity of the travelling waves.

Putting γ = surface tension of the soap solution (and there are two surfaces to the film), b and t = breadth and thickness of the film, ρ = density of the soap solution (nearly 1 gm. cm.^{3}) we get

$$v = \sqrt{\frac{2\gamma b}{\rho b t}} = \sqrt{\frac{2\gamma}{\rho t}}$$

Also if n = the frequency and λ = the wave-length

$$v = n\lambda$$

$$\therefore \quad n\lambda = \sqrt{\frac{2\gamma}{\rho t}}$$

$$\text{and} \qquad t = \frac{2\gamma}{n^2\lambda^2\rho}$$

In some instances of my own (see later) $n = 2 \times 64$ per second and λ was about 2 cm.; γ is about 27 dynes/cm.

$$\therefore \quad t = \frac{2 \times 27}{2^2 \times 64^2 \times 2^2 \times 1} = \frac{1}{1200} \text{ cm.}$$

which is of the right order of magnitude.

In the early days of the late war a mathematical colleague working on the propagation of electromagnetic waves along the surfaces of metals asked me to study the passage of transverse waves along long and narrow soap films. He thought that at certain frequencies there might be similarities between the two phenomena, and also that if it were possible to observe a transverse wave spreading around the corner of an L-shaped film, the appearance of the wave-fronts might help him in his solution of the electrical problem. After seeing what I could show him he appeared satisfied, but as his work was very much "Hush, Hush" I never learnt much about his own problem. Possibly it was connected with the important subject of "wave-guides" in radar.

In the experiments to be described the soap films were formed on slotted brass frames (Fig. 1). In Fig. 1(*a*) the slot was 17 cm. × 4 cm In Fig. 1(*b*) the slot was 30 cm. × 1 cm. In Fig. 1(*c*) the two separate brass right-angular pieces could be clamped together by the crossbars at the ends *A* and *C* to give an L-shaped film of adjustable length and width. The edges of the brass next the slot were bevelled to keep the film to the upper surface of the brass. The figure also shows little pieces of mica, *M*, M_1, which when wetted served to limit the extent of the film and when placed as at *M* in Fig. 1(*c*) acted as a reflector of the wave train which after travelling down *AB* was sent down *BC*.

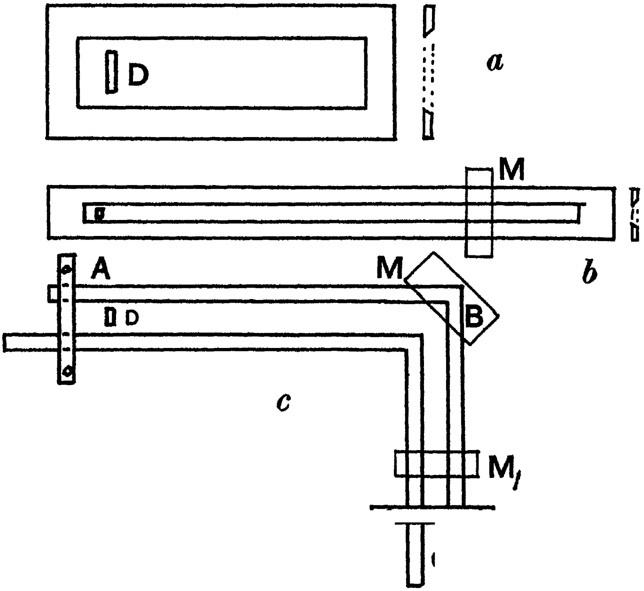


FIGURE 1. (*a*) The shorter broader slot. The sectional end view shows the bevelled edge of the slot. The dipper is shown at *D*.

(*b*) The longer narrow slot. *M* is a piece of wetted mica which serves, by its position, to limit the length of the film in use.

(*c*) The L-shaped rods used to give L-shaped films *ABC* of different lengths and breadths. The cross pieces at the ends are screw-clamps. *M* is a piece of mica placed at *B* to act as a reflector. M_1 is a second strip of mica to limit the length of the film along *BC*.

As each frame was used it was placed horizontally and the slot filled with the soap film. The dipper *D*, the source of the waves, was a small mica sheet attached either to a spring actuated by an A-C electromagnet or to the prong of an electromagnetically excited tuning fork. The dipper vibrated vertically, its amplitude was adjustable. Sometimes to give it a better grip of the soap film it was loaded with plasticene. The A-C electromagnet (or the tuning fork) was placed on a separate bench so that its vibrations were not communicated directly to the frame.

The observations were visual, and to help in the "seeing" a brightly illuminated pane of frosted glass was placed behind the soap film and the reflection of this glass in the surface of the film was observed at a large angle of incidence, sometimes at nearly grazing incidence.

Boys' soap solution (as supplied by Griffin and Tatlock) was used; it was applied to the frame by means of a sheet of celluloid.

The A-C frequency was about 60 to 64 per second. If standing waves are produced they are easily seen by the unaided eye by the varying reflectivity of the crests, troughs, and slopes.

Travelling waves will be too rapid for the unaided eye to note their pattern. In this case a Rayleigh phonic wheel (or synchronous motor) made by Pye and Company was used. The iron wheel of the motor has thirty teeth and the disc thirty slots. It was inserted in the A-C circuit or the fork circuit and once spun at the right speed is maintained at that speed keeping time with the A-C or fork. With this help the eye looks through the passing slots and the travelling waves now give a stationary pattern; and their wave-length can be estimated by counting the number in a given length. To make doubly sure that it was travelling waves that were under observation the wheel was accelerated by hand when the waves would apparently travel backward towards the source, or if slightly retarded by a little friction would apparently travel outwards from the source.

The film often breaks and if this happens when things look interesting a great demand is made on one's patience.

As mentioned above the frequency of the A-C supply was about 60 per second; the dipper frequency is twice this. The tuning fork used had frequencies 30, 120, 200 per second. A radio instrument known as a power-pack was also employed to operate the electromagnet. This has the advantage that the frequency can be changed at will and the varying effects observed.

The behaviour of the film was unpredictable. Sometimes the film just pulsated up and down, sometimes the liquid in the film appeared to be drawn towards the dipper and drops of liquid danced behind the dipper (in the confined space there—see Fig. 1). Sometimes the standing waves appeared at once—their wave-length being about a centimetre or so. Sometimes nothing could be seen unless the aid of the phonic wheel was invoked and then sometimes—not always—a set of travelling waves would be seen also of a wave-length of about a centimetre.

A few details with the frames are here given:

1942, July 15 to 31

With the long narrow trough and 120 and 30 forks only a local disturbance.

With the wide trough and the 120-fork surface drift as mentioned above.

With a 200-fork—travelling waves, λ between 1 and 2 cm.

With the power-pack operating the dipper and no plasticene used;

Frequency = 236. Travelling waves, 7 wave-lengths in 6½ cms.
Frequency = 200. The same but not so strong.
Frequency = 150. The same but λ greater.
Frequency = 100. The same but waves weaker.
Frequency = 56. The same but waves weaker still.
Frequency = 30. No waves, just a tremulous motion of film.

These were repeated with about the same results; then tried

Frequency = 400. No results seen.

1942, August 20

Using a soap solution which had been standing in an open dish for about a month and which had got quite viscous, I obtained stationary waves at times and travelling waves at other times with both frames of Figs. 1(*a*) and 1(*b*). Frequency, 110 per second. The interesting feature with the stationary waves was that as viewed at grazing incidence almost along the length of the slot the dark bands were seen to be not at right angles to the length of the slot but inclined to the length at about 45° (Fig. 2).

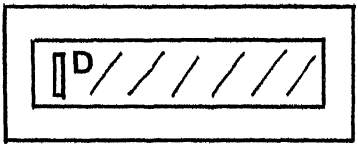


FIGURE 2.—A representation of the oblique dark bands seen when an extended bright source of light is viewed by reflection in the film at nearly grazing incidence.

There were 6 waves in about 9 cms. of length. With a fresh solution from the bottle only moving waves could be obtained.

More noteworthy still was that once with an A-C frequency of 56, travelling waves only were observed and the dark bands observed at merely grazing incidence formed a series of X's along the slot (Fig. 3),

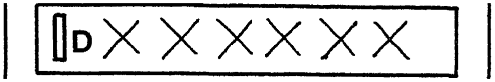


FIGURE 3.—A representation of the crossing dark bands (X-shaped) seen sometimes. Lighting and viewing conditions as in Fig. 2.

about 5 X's in a length of 9 cms. These X's got my mathematical colleague really excited. Are they evidence of waves travelling to the right and being reflected at the end and sent back?

1942, September 14 to 18

Using the L-shaped frame, Fig. 1(*c*), with the *AB* slot 2 cm. wide and the *BC* slot $1\frac{1}{2}$ cm. wide most interesting results were obtained. If the mica slip were not in position at *M*, waves only were seen in *AB*. With the slip present and its edge inclined at 45° as shown the waves penetrated along *BC* as well. In one case five standing waves were observed in a 7 cm. length of both *AB* and *BC*, the *BC* waves disappearing if the mica were pushed aside (Fig. 4). Also the dark bands were inclined to the length of the slot as mentioned above. Travelling waves (i.e., waves observed only with the phonic wheel) were also sometimes seen.

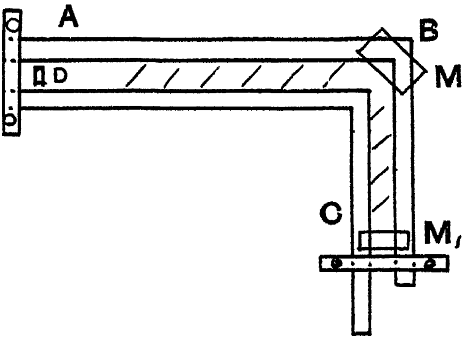


FIGURE 4.—A representation of the obliquely situated dark bands seen in the L-shaped film. The dark bands in *BC* are evidence that the waves are reflected by the edge of the mica slip at *M*.

Unfortunately, I have no photographic record of these phenomena. Possibly the variations in the phenomena are due to changes in the natural frequency of the film caused by changes in film-thickness. It is hoped to repeat some of the observations when more time is available.

TRANSACTIONS OF

THE ROYAL SOCIETY OF CANADA

SECTION III

CHEMICAL, MATHEMATICAL, AND PHYSICAL SCIENCES

THIRD SERIES -VOLUME XLII SECTION III

MAY, 1948

56670

OTTAWA
THE ROYAL SOCIETY OF CANADA
1948

TABLE OF CONTENTS

* * *

Photosensitization. Presidential Address, by E. W. R. STEACIE, F.R.S.C. 1

On the Minimum Value of the Riemann-Roch Expression for Order-Bases in the Large. By S. BEATTY, F.R.S.C. . . . 11

On the Number of Conditions to Apply to a Function R(Z, U) *to Build it on an Assigned Local Order-Basis* (τ). By S. BEATTY, F.R.S.C. 15

Illumination for Concealment of Ships at Night: Some Technical Considerations. By E. GODFREY BURR 19

The Determination of the International Volt to within Twenty Microvolts by the Application of Ageing Corrections to Old Normal Weston Standard Cells. By A. HENRIKSON and A. NORMAN SHAW, F.R.S.C. 37

On Perron's Extension of the Euler-Knopp Summation Method. By M. S. MACPHAIL 43

The Stability of Gas Turbine Combustors. By DONALD L. MORDELL 51

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLII : SERIES III : MAY, 1948
SECTION THREE

PRESIDENTIAL ADDRESS

Photosensitization

By E. W. R. STEACIE, F.R.S.C

IN order that a photochemical reaction may occur it is necessary in the first place that the magnitude of the quantum of the incident light be large enough, and secondly that the light be absorbed. In many cases substances have dissociation energies corresponding to frequencies in a convenient region of the spectrum, but are transparent down to the vacuum ultra-violet where photochemical experiments are difficult to carry out. Thus hydrogen has a dissociation energy of 102.7 kcal. corresponding to the energy of a quantum at 2776A, but the continum in its absorption spectrum does not begin till 849A.

If we add to the hydrogen a substance which absorbs energy of some wave-length less than 2776A, and which will on collision transfer this energy by some mechanism to a hydrogen molecule, then it will be possible to effect the dissociation by means of radiation in an accessible spectral region. Such a process is called photosensitization. Mercury vapour has been the most widely used sensitizer, and we will discuss it first as an example.

RESONANCE RADIATION

Fig. 1 gives an energy level diagram for the mercury atom. If a mercury atom in the ground 6^1S_0 state absorbs λ 2537 it will be raised to the 6^3P_1 state, which has an average life of 1.08×10^{-7} sec. If the atom is left to itself it will drop back to the ground state and λ2537 will appear again as fluorescence, or so-called resonance radiation. If, however, the excited mercury atom makes an appropriate collision with a foreign gas molecule, it may be *quenched* and lose its energy without an optical transition occurring. The extent of such quenching will depend in the first place on the number of collisions with foreign gas molecules, i.e., on the pressure of the foreign gas. In the second place it will depend on the efficiency of such collisions. This is usually expressed in terms of a cross-section for quenching. Such cross-sections vary widely, from 0.08×10^{-16} cm.^{2} for CH_4 to 60×10^{-16} cm.^{2} for C_6H_6. (18, 21, 37.)

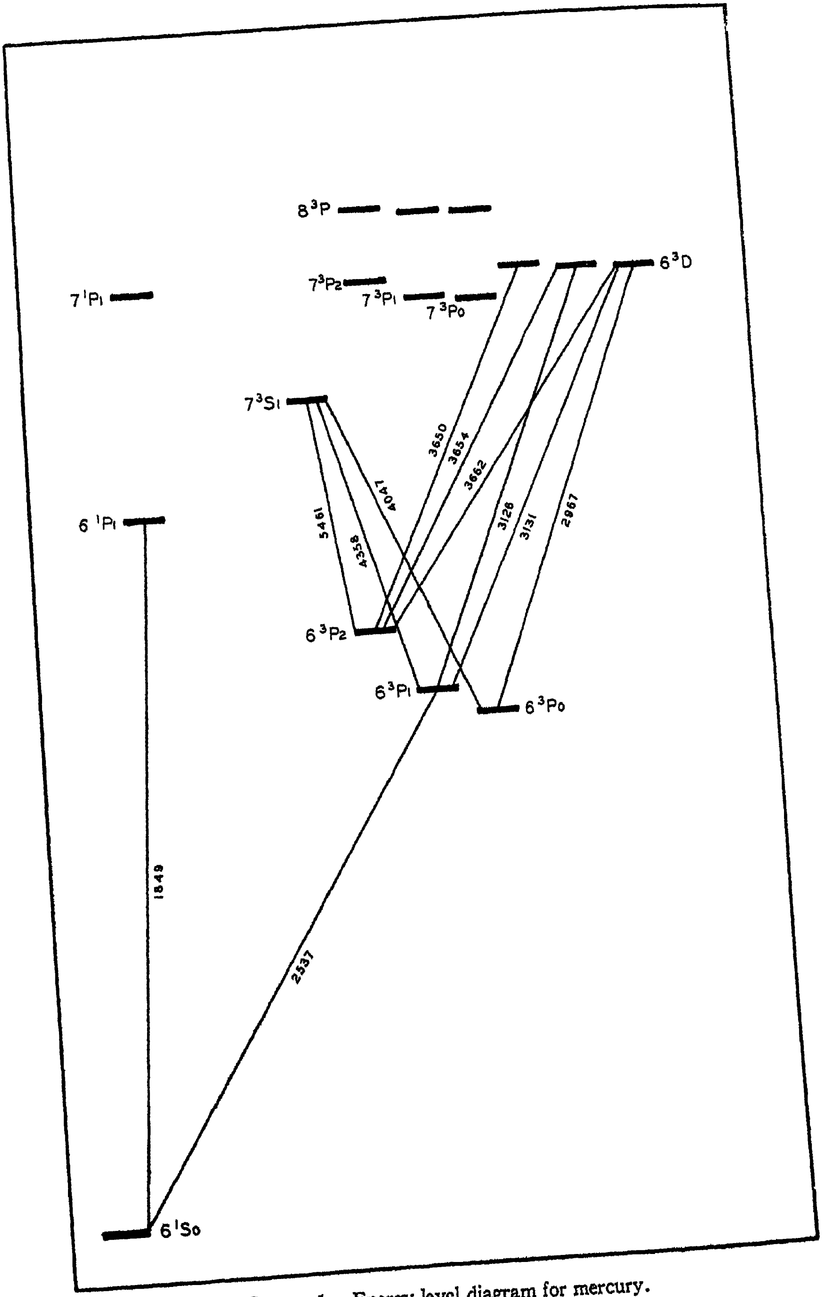


FIGURE 1.—Energy level diagram for mercury.

THE QUENCHING PROCESS

An excited mercury atom by collision with a molecule of a foreign gas may depart from the 6^3P_1 state with a loss of energy in three fundamentally different ways:

1. *The mercury atom is transferred to the metastable* 6^3P_0 *state.* The excitation energies of the states are·

$$6^3P_1 = 4.86 \text{ eV, or } 112\ 20 \text{ kcal.}$$
$$6^3P_0 = 4.64 \text{ eV, or } 107.17 \text{ kcal.}$$

Hence, in dropping to the metastable state, the energy transferred to the foreign gas molecule cannot exceed 0.22 eV or 5 kcal. This is far too small for chemical reaction or an electronic transition, and the energy will ultimately merely appear as heat. That quenching does occur in this way in certain cases has been proved by Wood and others in the case of nitrogen.

2. *Quenching by an electronic transition.* In this type of process we have:

$$Hg(6^3P_1) + RH \rightarrow Hg(6^1S_0) + RH^*, \qquad (1)$$

where RH^* represents an electronically excited molecule of a foreign gas. In this case the amount of energy which may be transferred is large, up to 112.2 kcal., and is sufficient to break most chemical bonds if it can be efficiently used. There is one important restriction on this type of process which will be discussed later.

The excited molecule formed in (1) may, in general, either decompose or be deactivated, i.e.

$$RH^* + RH \rightarrow 2RH \qquad (2)$$

or

$$RH^* \rightarrow R + H. \qquad (3)$$

Reaction (3) involves an internal conversion of the electronic energy of the excited molecule to vibrational energy, and if this is effective, decomposition of the molecule will follow.

From a kinetic standpoint the life of the excited molecule RH^* is of importance, since if the life is so short that deactivation never occurs, then (1) followed by (3) will be indistinguishable from the one-step process

$$Hg(^3P_1) + RH \rightarrow Hg(^1S_0) + R + H. \qquad (4)$$

3. *Quenching by dissociation.* A third mechanism of quenching involves dissociation of the quenching molecule in one step during the quenching process. This may occur either directly or via compound formation with the sensitizer, i.e., by (4), or by

$$Hg(^3P_1) + RH \rightarrow HgH + R \qquad (5)$$

followed by

$$HgH \rightarrow Hg + H. \qquad (6)$$

HgH is very weakly bound, its heat of formation (from the atoms) being only 8.5 kcal. It therefore dissociates rapidly even at room temperature. As pointed out above, reaction (1) can really be regarded as a special case of the excited molecule mechanism in which the excited molecule has a very short life.

There has been considerable controversy over the question of the participation of HgH in mercury sensitized reactions. It has more recently been found by Olsen, however, that it is not possible to produce resonance excitation of the HgH bands in mixtures of hydrogen and mercury. (20.) He concluded that HgH is not present in its normal state, but is formed in excited states by secondary processes involving hydrogen atoms. It is, however, possible that HgH is formed by reaction (5), but that it has sufficient energy to enable its almost immediate dissociation. In other words, if (6) follows (5) rapidly enough the over-all process becomes indistinguishable from (4).

The intermediate formation of HgH may occur also in the case of hydrocarbons, e.g.

$$Hg(^3P_1) + C_2H_6 \rightarrow HgH + C_2H_5. \qquad (7)$$

This would merit investigation. The effect is of some importance since if HgH is formed as an intermediate it is possible for a bond to be broken in the reaction whose strength is equal to

$$\begin{aligned}&(\text{excitation energy of } Hg(^3P_1) + \text{heat of formation of } HgH)\\ &= 112.2 + 8.5 = 120.7 \text{ kcal.}\end{aligned}$$

A COMPARISON OF ZINC, CADMIUM, AND MERCURY

By using other metals as sensitizers it is possible to vary the energy input to the reacting molecule. Thus with mercury, cadmium, and zinc the following excitation energies are involved:

State	*Line*	*Excitation energy*
$Hg(6^1P_1)$	1849A	153.9 kcal.
$Hg(6^3P_1)$	2537	112.2
$Cd(5^1P_1)$	2288	124.4
$Cd(5^3P_1)$	3261	87.3
$Zn(4^1P_1)$	2139	133.4
$Zn(4^3P_1)$	3076	92.5

In the case of cadmium (and probably zinc), however, there is one important difference. As we have seen, Olsen's work makes it probable that with mercury-hydrogen mixtures the primary step is a direct split into two hydrogen atoms, rather than an intermediate formation of mercury hydride. However, in the case of cadmium-hydrogen mixtures illuminated with the resonance line at 3261A, Bender and Olsen have obtained strong resonance excitation of CdH bands. (1, 20.) This indicates that CdH is formed in the normal state and suggests that the primary step is

$$Cd(5^3P_1) + H_2 \rightarrow CdH + H. \qquad (8)$$

It is therefore not at all certain *a priori* whether with reactions involving a C-H bond split in a hydrocarbon or other organic molecule we have in a given case

$$M(^3P_1) + RH \rightarrow M(^1S_0) + R + H \qquad (9)$$

or

$$M(^3P_1) + RH \rightarrow MH + R. \qquad (10)$$

The heat of formation of CdH is 15.5 kcal. and of ZnH 23.1 kcal. C-H bonds in most organic molecules have strengths in the neighbourhood of 95-103 kcal. Hence, if (9) is the mechanism, $Cd(5^3P_1)$ and $Zn(4^3P_1)$ atoms would be incapable of causing a C-H split. However, if the reaction proceeds via (10), bonds can be broken with strengths up to 87.3 + 15.5 = 102.8 kcal. and 92.5 + 23.1 = 115.6 kcal. respectively.

In the case of propane-cadmium mixtures illuminated with the cadmium resonance line at 3261A (thus giving rise to Cd (5^3P_1) atoms) the resonance excitation of CdH bands has been observed and there is thus no question that the quenching process involves (10). (27.)

SPIN CONSERVATION IN THE QUENCHING PROCESS

The processes occurring with hydrogen and a variety of hydrocarbons have recently been reviewed by Laidler and by the author. (11, 22.) In this section we will present views which are largely due to Laidler. Let us consider the quenching process in more detail, and for simplicity start with hydrogen.

With singlet metal atoms there are two possible primary processes:

$$M(^1P) + H_2(^1\Sigma g^+) \rightarrow M(^1S) + H_2(^1\Sigma)^* \qquad (11)$$

where $H_2(^1\Sigma)^*$ denotes a hydrogen molecule in its normal $^1\Sigma$ electronic state, but containing a large amount of vibrational energy. If formed, it will undoubtedly usually dissociate,

$$H_2(^1\Sigma)^* \rightarrow 2H(^2S). \qquad (12)$$

The second possibility is

$$M(^1P) + H_2(^1\Sigma) \rightarrow M(^2S) + 2H(^2S). \qquad (13)$$

It will be seen that the spin-conservation rule is obeyed in both cases.

With triplet atoms the possible processes are

$$M(^3P) + H_2(^1\Sigma) \rightarrow M(^1S) + H_2(^1\Sigma)^* \qquad (14)$$

or

$$\rightarrow M(^1S) + 2H(^2S). \qquad (15)$$

In this case (15) is satisfactory from a multiplicity point of view, but (14) is ruled out. Actually, (15) undoubtedly proceeds via the intermediate formation of a complex, viz.

$$M(^3P) + H_2(^1\Sigma) \rightarrow MH_2(^3\Sigma)$$
$$\rightarrow M(^1S) + 2H(^2S). \qquad (16)$$

It follows that with singlet metal atoms either the formation of an excited molecule or a direct split to atoms is possible. With triplet atoms, on the other hand, only the direct split to atoms is possible.

In the case of hydrocarbon molecules the ground state is undoubtedly $^1\Sigma$, and the situation is thus analogous to that of hydrogen. We may therefore consider the reactions from a similar point of view. Thus with singlet atoms we have:

$$M(^1P) + RH(^1\Sigma) \rightarrow M(^1S) + RH(^1\Sigma)^* \qquad (17)$$

or

$$\rightarrow M(^1S) + R(^2\Sigma) + H(^2S). \qquad (18)$$

Either (17) or (18) is possible from a spin-conservation point of view.

With triplet atoms we have the possibilities

$$M(^3P) + RH(^1\Sigma) \rightarrow M(^1S) + RH(^1\Sigma)^* \qquad (19)$$
$$\rightarrow M(^1S) + R(^2\Sigma) + H(^2S) \qquad (20)$$
$$\rightarrow M(^1S) + RH(^3\Sigma). \qquad (21)$$

Reaction (19) is ruled out on a multiplicity basis. Hence the only way in which an excited molecule can be formed is by (21), and this presupposes the existence of a low-lying triplet state.

It should be noted that in the case of (20) it is immaterial whether the intermediate formation of a metal hydride occurs or not, i.e., we can equally well have

$$M(^3P) + RH(^1\Sigma) \rightarrow MH(^2\Sigma) + R(^2\Sigma) \qquad (22)$$

followed by

$$MH(^2\Sigma) \rightarrow M(^1S) + H(^2S). \qquad (23)$$

REACTIONS OF HYDROCARBONS

Investigations have been made recently of the reactions of a variety of paraffins and olefins, as follows:

With $Hg(^3P_1)$
ethane (3, 23, 29, 30, 31), propane (2, 24), n-butane (10), iso-butane (3), neopentane (4), ethylene (12, 13, 16), propylene (6), 1-butene (8), 2-butene (8), isobutene (7), acetylene (15), butadiene (5, 36), and isoprene (6);

With Cd($^3P_1$)
ethane (33), propane (28), ethylene (14, 32, 34, 35), propylene (14), 1-butene (14), 2-butene (14), and acetylene (15);

With Cd($^1P_1$)
ethylene (25), and acetylene (15);

With Zn($^3P_1$)
ethylene (9);

With Zn($^1P_1$)
ethylene (9).

Certain general conclusions may be drawn from the results, which will be reviewed in the light of the foregoing discussion.

1. *Singlet atoms with paraffins.* No investigations have been made.

2. *Triplet atoms with paraffins.* The paraffins have small quenching cross-sections. There is no evidence of the deactivation of an excited molecule, since the rate of reaction rises rather than falls with increasing pressure. The experimental evidence very strongly favours an initial C-H bond split, and in the case of ethane with Hg($^3P_1$) hydrogen atoms have definitely been shown to be produced. (4.) The reactions therefore presumably proceed by (20) or (22) followed by (23). In the case of Cd($^3P_1$) a C-H bond split can only occur via the intermediate formation of CdH for energetic reasons. (27.) In the case of propane with Cd($^3P_1$), CdH has been shown to be present by the resonance excitation of its band spectrum. Reaction (19) is presumably absent because of the lack of a suitable triplet excited state with the paraffins.

The subsequent reactions of the atoms and radicals produced in the primary step are fairly well established. The reaction of ethane with Hg($^3P_1$) may be satisfactorily accounted for by the mechanism:

$$Hg(^3P_1) + C_2H_6 \rightarrow Hg(^1S_0) + C_2H_5 + H \qquad (24)$$

$$H + C_2H_6 \rightarrow C_2H_5 + H_2 \qquad (25)$$

$$H + C_2H_5 \rightarrow C_2H_6^* \qquad (26)$$

$$C_2H_6^* \rightarrow 2CH_3 \qquad (27)$$

$$C_2H_6^* + C_2H_6 \rightarrow 2C_2H_6. \qquad (28)$$

3. *Singlet atoms with olefins.* No quenching data exist, since the life of the excited singlet states is so short that imprisonment phenomena make measurements impossible. There are, perhaps, indications that the quenching cross-sections may be small. There is no evidence of deactivation processes, and the mechanism appears to involve mainly a C-H split, i.e. reaction (18). In the case of ethylene, the main products are those to be expected from an atom and radical mechanism.

A small amount of acetylene is also formed, indicating that we also have, but to a very minor extent,

$$Cd(^1P_1) + C_2H_4 \rightarrow Cd(^1S_0) + C_2H_4^1 \qquad (29)$$

$$C_2H_4^1 \rightarrow C_2H_2 + H_2. \qquad (30)$$

4. *Triplet atoms with olefins.* This is by far the most thoroughly investigated case. The values of the quenching cross-sections are in all cases very large. The indications are that reactions (20) and (21) both occur, but that in general quenching by (21) is much more important than by (20).

In some cases, e.g. ethylene and $Cd(^3P_1)$, quenching by (21) to form an excited molecule is very efficient, but there is insufficient energy for subsequent reaction of the excited molecule.

There is strong evidence for deactivation of an excited molecule in most cases. The indications are that the excited molecule may subsequently react by a C-H bond split, or by the elimination of a hydrogen molecule to form an acetylenic compound. The possibilities thus are:

$$M(^3P_1) + RH \rightarrow M(^1S_0) + RH^*$$

$$RH^* \rightarrow R + H$$

or

$$\rightarrow H_2 + \text{acetylenic compound}$$

or

$$\rightarrow \text{no reaction, if the energy is insufficient.}$$

As an example of the subsequent reactions we may consider the case of ethylene, which has been the most thoroughly investigated. (12, 13.)

With triplet mercury atoms the main products of the reaction at room temperature are hydrogen, acetylene, butane, and butene. There is an initial pressure rise, followed by a decrease. At low pressures the initial pressure rise corresponds with the amount of hydrogen and acetylene formed. The rate of the reaction decreases with increasing ethylene pressure, pointing to a deactivation process. The initial steps in the mechanism are well established as:

$$Hg(^3P_1) + C_2H_4 \rightarrow C_2H_4^* + Hg(^1S_0) \qquad (31)$$

$$C_2H_4^* + C_2H_4 \rightarrow 2C_2H_4 \qquad (32)$$

$$C_2H_4^* \rightarrow C_2H_2 + H_2. \qquad (33)$$

As the reaction progresses hydrogen builds up and the reaction

$$Hg(^3P_1) + H_2 \rightarrow 2H + Hg(^1S_0) \qquad (34)$$

comes into play. Subsequent reactions of H-atoms with ethylene lead to the formation of C_4 hydrocarbons, and to the subsequent decrease in pressure. There are also indications that the alternative primary process

$$Hg(^3P_1) + C_2H_4 \rightarrow C_2H_3 + H + Hg(^1S_0) \qquad (35)$$

comes into play to a small extent.

At higher temperatures the initial pressure rise disappears. It is restored, however, by the addition of nitric oxide which acts as an inhibitor of chain processes. The rate of ethylene disappearance also increases at high temperatures. There seems to be no doubt that at the higher temperatures chains are initiated by H-atoms and vinyl radicals from (35).

With triplet cadmium atoms quenching by ethylene is very strong, but very little reaction occurs. (14, 26, 34.) Apparently the process

$$Cd(^3P_1) + C_2H_4 \rightarrow C_2H_4^* + Cd(^1S_0) \qquad (36)$$

takes place with high efficiency, but because of energy restrictions the subsequent reactions of $C_2H_4^*$ do not occur. There is no evidence for deactivation processes, and it seems likely that what reaction does occur proceeds via

$$Cd(^3P_1) + C_2H_4 \rightarrow CdH + C_2H_3 \qquad (37)$$

followed by

$$CdH \rightarrow Cd(^1S_0) + H. \qquad (38)$$

If the reaction goes in this way 102.8 kcal. are available to break the C-H bond, which seems ample.

REFERENCES

1. BENDER, P. (1930). Phys. Rev. **36**: 1535.
2. DARWENT, B. DE B. and STEACIE, E. W. R. (1945). J. Chem. Phys. **13**: 563.
3. DARWENT, B. DE B. and WINKLER, C. A. (1945). J. Phys. Coll. Chem. **49**: 150.
4. DARWENT, B. DE B. and STEACIE, E. W. R. (1948). J. Chem. Phys. **16**: 381.
5. GUNNING, H. E. and STEACIE, E. W. R. (1944). J. Chem. Phys. **12**: 484.
6. GUNNING, H. E. and STEACIE, E. W. R. (1946). J. Chem. Phys. **14**: 57.
7. GUNNING, H. E. and STEACIE, E. W. R. (1946). J. Chem. Phys. **14**: 544.
8. GUNNING, H. E. and STEACIE, E. W. R. (1946). J. Chem. Phys. **14**: 581.
9. HABEEB, H., LEROY, D. J., and STEACIE, E. W. R. (1942). J. Chem. Phys. **10**: 261.
10. HAY, A. W. and WINKLER, C. A. (1943). Can. J. Res. B21: 149.
11. LAIDLER, K. J. (1947), J. Chem. Phys. **15**: 712.
12. LEROY, D. J. and STEACIE, E. W. R. (1941). J. Chem. Phys. **9**: 829.
13. LEROY, D. J. and STEACIE, E. W. R. (1942). J. Chem. Phys. **10**: 676.
14. LEROY, D. J. and STEACIE, E. W. R. (1943). J. Chem. Phys. **11**: 683.
15. LEROY, D. J. and STEACIE, E. W. R. (1944). J. Chem. Phys. **12**: 117.
16. LINE, D. and LEROY, D. J. (1945). J. Chem. Phys. **13**: 307.
17. LIPSON, H. C. and MITCHELL, A. C. G. (1935). Phys. Rev. **48**: 625.
18. MITCHELL, A. C. G. and ZEMANSKY, M. (1934). Resonance radiation and excited atoms. New York: Macmillan.

19. Norrish, R. G. W. and Smith, W. McF. (1940). Proc. Roy. Soc. A**176**: 295.
20. Olsen, L. O. (1938). J. Chem. Phys. **6**: 307.
21. Steacie, E. W. R. (1940). Can. J. Res. B**18**: 44.
22. Steacie, E. W. R. (1946). Atomic and free radical reactions. New York: Reinhold.
23. Steacie, E. W. R. and Cunningham, R. L. (1940). J. Chem. Phys. **8**: 800.
24. Steacie, E. W. R. and Dewar, D. J. (1940). J. Chem. Phys. **8**: 571.
25. Steacie, E. W. R. and LeRoy, D. J. (1942). J. Chem. Phys. **10**: 22.
26. Steacie, E. W. R. and LeRoy, D. J. (1943). J. Chem. Phys. **11**: 164.
27. Steacie, E. W. R. and LeRoy, D. J. (1944). J. Chem. Phys. **12**: 34.
28. Steacie, E. W. R., LeRoy, D. J., and Potvin, R. (1941). J. Chem. Phys. **9**: 306.
29. Steacie, E. W. R. and Phillips, N. W. F. (1938). J. Chem. Phys. **6**: 179.
30. Steacie, E. W. R. and Phillips, N. W. F. (1938). Can. J. Res. B**16**: 303.
31. Steacie, E. W. R., Phillips, N. W. F., and Alexander, W. A. (1938). Can. J. Res. B**16**: 314.
32. Steacie, E. W. R. and Potvin, R. (1938). Can. J. Res. B**16**: 337.
33. Steacie, E. W. R. and Potvin, R. (1939). J. Chem. Phys. **7**: 782.
34. Steacie, E. W. R. and Potvin, R. (1940). Can. J. Res. B**18**: 47.
35. Taylor, H. S. and Bates, J. R. (1927). J. Amer. Chem. Soc. **49**: 2438.
36. Volman, D. H. (1946). J. Chem. Phys. **14**: 467.
37. Zemansky, M. (1930). Phys. Rev. **36**: 919.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLII : SERIES III : MAY, 1948
SECTION THREE

On the Minimum Value of the Riemann-Roch Expression for Order-Bases in the Large

By S. BEATTY, F.R.S.C.

LET the fundamental equation be

$$F(Z, U) \equiv U^N + U^{N-1}F_1(Z) + \ldots + F_N(Z) = 0,$$

where each $F_i(Z)$ is the quotient of two polynomials in Z, with coefficients taken from the field of complex numbers. It may be supposed that F possesses ρ irreducible factors, all different and all of the same general character as F itself.

In the vicinity of a point $Z = a$ (or $Z = \infty$), the N expansions of U in powers of $Z - a$ $\left(\text{or } \frac{1}{Z}\right)$, fall into r cycles containing

$$\nu_1, \nu_2, \ldots, \nu_r \tag{1}$$

expansions, where the expansions of a single cycle are all conjugate to one another, in the sense that they permute among themselves when roots of unity appearing in them are replaced by their conjugates.

A rational function $R(Z, U)$ involving constants taken from the field of complex numbers has order-numbers in the vicinity of a point $Z = a$ (or $Z = \infty$), which are multiples of

$$\frac{1}{\nu_1}, \frac{1}{\nu_2}, \ldots, \frac{1}{\nu_r} \tag{2}$$

for branches belonging to the corresponding cycles of (1). These multiples are all finite, except in the case of rational functions R which in reduced form are divisible by one or more of the ρ irreducible factors of F. In particular, the rational function $\frac{\partial F}{\partial U}$ has order-numbers

$$\mu_1 = \frac{m_1}{\nu_1}, \mu_2 = \frac{m_2}{\nu_2}, \ldots, \mu_r = \frac{m_r}{\nu_r},$$

which are all finite. Any set of finite multiples of (2) of the form

$$\tau_1 = \frac{n_1}{\nu_1}, \tau_2 = \frac{n_2}{\nu_2}, \ldots, \tau_r = \frac{n_r}{\nu_r}.$$

is called a local order-basis corresponding to the point $Z = a$ (or $Z = \infty$), and may be denoted by $(\tau)_a$ (or $(\tau)_\infty$). An order-basis in the large is made up of local order-bases, one for each point, such that all but a finite number of the multiples involved are zero. Such an order-basis in the large is denoted by $((\tau))$. Just as a rational function R which in reduced form is not divisible by any of the ρ irreducible factors of F supplies a local order-basis $(\tau)_a$ (or $(\tau)_\infty$), so also does it supply an order-basis in the large $((\tau))$, and this has the further property that

$$\sum\sum \tau\nu = 0\,.$$

Order-bases in the large $((\tau))$ and $((\tau))$ are complementary, provided each sum $\tau + \bar{\tau}$ of associated order-numbers taken from local order-bases is $\mu - 1 + \frac{1}{\nu}$ for finite cycles but $\mu + 1 + \frac{1}{\nu}$ for cycles at $Z = \infty$.

When an order-basis in the large $((\tau))$ is assigned, a rational function R is said to be built on it, if the order-basis in the large supplied by R has all its order-numbers equal to or greater than the corresponding order-numbers of $((\tau))$.

The number of linearly independent functions R built on an order-basis in the large $((\tau))$ is denoted by $N((\tau))$. The Riemann-Roch Theorem is contained in the formula

$$N((\tau)) + \tfrac{1}{2}\sum\sum \tau\nu = N((\bar{\tau})) + \tfrac{1}{2}\sum\sum \bar{\tau}\nu\,.$$

In other words, if the Riemann-Roch expression for an order-basis in the large $((\tau))$ is denoted by

$$N((\tau)) + \tfrac{1}{2}\sum\sum \tau\nu\,,$$

the Riemann-Roch Theorem merely asserts that the Riemann-Roch expressions for complementary order-bases in the large $((\tau))$ and $((\bar{\tau}))$ are equal.

It is the purpose of the present note to find the minimum value of all Riemann-Roch expressions corresponding to all order-bases in the large $((\tau))$ and to characterize the particular order-bases whose Riemann-Roch expressions take this minimum value.

It is evident that the Riemann-Roch expressions have a minimum. For the Riemann-Roch expression for $((\tau))$, being the same as that for $((\tau))$, is given by

$$\frac{N((\tau)) + N((\bar{\tau}))}{2} + \tfrac{1}{4}\sum\sum(\tau + \bar{\tau})\nu\,.$$

The first of these items is zero or more. The second is $\frac{1}{2}(\rho - p)$, as appears from applying the Riemann-Roch Theorem to complementary order-bases in the large $((0))$ and $((\tau + \bar{\tau}))$ and observing that

$$N((0)) = \rho,\; N((\tau + \bar{\tau})) = p\,.$$

It follows, therefore, that every Riemann-Roch expression is equal to or greater than $\frac{1}{2}(\rho - p)$.

We shall now show that the minimum value sought is precisely $\frac{1}{2}(\rho - p)$ and that it arises through the circumstance that

$$N((\tau)) = N((\bar{\tau})) = 0\,.$$

Suppose that $((\tau))$ and $((\bar{\tau}))$ are complementary order-bases in the large whose Riemann-Roch expressions take the minimum value which we are seeking to determine. Let now $((\tau'))$ be an order-basis in the large, identical with $((\tau))$, except that for one cycle of ν expansions it drops below $((\tau))$ by the least possible amount, namely $\dfrac{1}{\nu}$. The order-basis in the large complementary to $((\tau'))$ will be denoted by $((\bar{\tau}'))$. It is identical with $((\bar{\tau}))$, except that for the particular cycle in question it exceeds $((\bar{\tau}))$ by the least possible amount $\dfrac{1}{\nu}$. It is plain that the Riemann-Roch expression for $((\tau'))$ exceeds that for $((\tau))$ by $\frac{1}{2}$, since it must differ from it by $\frac{1}{2}$ but cannot be less than it by $\frac{1}{2}$. This means, therefore, that

$$N((\bar{\tau}')) = N((\bar{\tau}))\,.$$

In other words, the general function built on the order-basis in the large $((\bar{\tau}))$ has for every cycle an order beyond the order-basis on which it is built. This can only mean that $N((\bar{\tau})) = 0$, since 0 is the only function built on $((\bar{\tau}))$. In like manner, it follows that $N((\tau)) = 0$. This implies that the minimum value of the Riemann-Roch expression for order-bases in the large $((\tau))$ is $\frac{1}{2}(\rho - p)$ and that it arises only in the case of those complementary order-bases in the large $((\tau))$ and $((\bar{\tau}))$ for which

$$N((\tau)) = N((\bar{\tau})) = 0\,.$$

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLII : SERIES III : MAY, 1948
SECTION THREE

On the Number of Conditions to Apply to a Function $R(Z, U)$ to Build it on an Assigned Local Order-Basis (τ)

By S. BEATTY, F.R.S.C.

WHERE $F(Z, U) = 0$, of degree N in U, is the fundamental algebraic equation, and where (τ) denotes an assigned order-basis at $Z = 0$, the equivalent function-basis may be written in the form*

$$\frac{U^{N-1} + U^{N-2}H_{N-1N-2} + \ldots + H_{N-10}}{Z^{a_{N-1}}}, \ldots, \frac{U + H_{10}}{Z^{a_1}}, \frac{1}{Z^{a_0}},$$

in which the fundamental exponents (a) are rational integers, while the coefficients (H) are rational functions of Z. Indeed, if U is everywhere finite in the vicinity of $Z = 0$, it is known that

$$a_{N-1} \geqq \ldots \geqq a_1 \geqq a_0,$$

and each H_{pq} may be chosen a unique polynomial of degree less than $a_p - a_q$.

We propose to prove a theorem heretofore taken for granted. It may be stated as follows:

THEOREM. *Let* λ, *appearing in*

$$Q \equiv \frac{U^{N-1}Q_{N-1} + \ldots + UQ_1 + Q_0}{Z^{\lambda}},$$

be a sufficiently large rational integer so that Q *includes all rational functions* $R(Z, U)$ *built on an assigned order-basis* (τ) *at* $Z = 0$, *where*

$$Q_{N-1}, \ldots, Q_1, Q_0$$

are arbitrary polynomials in Z. *The number of conditions to apply to* Q *to build it on* (τ) *is exactly* $N\lambda - \Sigma a$.

We shall prove this first for the special case that U is everywhere finite in the vicinity of $Z = 0$. The conditions in question result from identifying Q with the function

$$P \equiv P_{N-1}\frac{U^{N-1} + U^{N-2}H_{N-1N-2} + \ldots + H_{N-10}}{Z^{a_{N-1}}} + \ldots$$
$$+ P_1\frac{U + H_{10}}{Z^{a_1}} + P_0\frac{1}{Z^{a_0}},$$

---

*Scripta Mathematica 5 (1938): 183-94, 239-45.

in which $P_{N-1}, \ldots, P_1, P_0$ are arbitrary polynomials in Z. The identification of Q with P is effected by insisting on the N relations:

$$Q_{N-1} = Z^{\lambda - a_{N-1}} P_{N-1},$$

$$Q_{N-2} = Z^{\lambda - a_{N-1}} P_{N-1} H_{N-1\,N-2} + Z^{\lambda - a_{N-2}} P_{N-2},$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$Q_0 = Z^{\lambda - a_{N-1}} P_{N-1} H_{N-1\,0} + Z^{\lambda - a_{N-2}} P_{N-2} H_{N-2\,0} + \ldots + Z^{\lambda - a_0} P_0.$$

According to the first of these relations, we must equate to 0 all coefficients of powers of Z less than $\lambda - a_{N-1}$ appearing in Q_{N-1}, and after that we must express all coefficients of higher powers of Z appearing in Q_{N-1} linearly in terms of arbitrary coefficients in P_{N-1}, all of which are involved in this system of linear equations. It is plain that the first $\lambda - a_{N-1}$ conditions are all effective, since they restrict the values of the coefficients specified in Q_{N-1}. But, all the remaining conditions are merely nominal, in the sense that they provide a scheme for naming the coefficients left over in Q_{N-1} in terms of arbitrary coefficients in P_{N-1}. Alternatively, this naming scheme could be regarded as operating in the opposite sense, namely from P_{N-1} to Q_{N-1}.

According to the second of these relations, we have to put equal to 0 all coefficients of powers of Z less than $\lambda - a_{N-1}$ appearing in Q_{N-2}, then express all coefficients of powers of Z less than $\lambda - a_{N-2}$ but not less than $\lambda - a_{N-1}$ appearing in Q_{N-2} linearly in terms of arbitrary coefficients in P_{N-1}, and finally express all coefficients of powers of Z not less than $\lambda - a_{N-2}$ appearing in Q_{N-2} linearly in terms of arbitrary coefficients in P_{N-1}, P_{N-2}, all the coefficients from the latter of these arbitrary polynomials being definitely involved in this system of linear equations. For the same reason as that stated in the case of the first relation, we have to count the $\lambda - a_{N-1}$ conditions in the first group as effective. The $a_{N-1} - a_{N-2}$ conditions in the second group must also be counted as effective, since they relate coefficients appearing in Q_{N-2} with coefficients appearing in Q_{N-1} through the medium of two schemes for naming these Q_{N-2} and Q_{N-1} coefficients in terms of those of P_{N-1}. The conditions in the third group are all nominal, seeing that they involve for the first time the coefficients of P_{N-2}. The total of new effective conditions implied in the second relation taken as succeeding the first is, therefore, $\lambda - a_{N-2}$.

It is clear by an analogous argument that the third relation is responsible for $\lambda - a_{N-3}$ new effective conditions, of which $\lambda - a_{N-1}$ arise from the third relation taken by itself, $a_{N-1} - a_{N-2}$ arise from the third relation taken after the first two, and $a_{N-2} - a_{N-3}$ arise from the third relation also taken after the first two. There are, of course, nominal conditions here also, which do not count, since they involve for the first time the coefficients of P_{N-3}.

Indeed, the argument continues to apply throughout the whole schedule of N relations. The number of conditions all told is, therefore,

$$(\lambda - a_{N-1}) + (\lambda - a_{N-2}) + \dots + (\lambda - a_0)\,,$$

and since from the manner of their derivation it is clear that they are all independent, their total is the number sought, namely $N\lambda - \Sigma a$. This completes the proof of the theorem in the special case under discussion.

It remains to show that the result continues to hold even if U is not everywhere finite in the vicinity of $Z = 0$. The functions Q and P will be taken as before. It may be, however, that the inequalities

$$a_{N-1} \geqq \dots \geqq a_1 \geqq a_0$$

no longer hold, and besides it may happen that the functions (*II*) are no longer polynomials. In any event, let a positive integer h be chosen so that $V = UZ^h$ is everywhere finite in the vicinity of $Z = 0$. It follows that

$$Q \equiv \frac{V^{N-1}Q_{N-1} + \dots + VZ^{(N-2)h}Q_1 + Z^{(N-1)h}Q_0}{Z^{\lambda + (N-1)h}}\,.$$

It appears that the exponent λ appearing in Q expressed in terms of U gives place to the corresponding exponent $\lambda + (N-1)h$ when Q is expressed in terms of V. Besides, the fundamental exponents

$$a_{N-1}, \dots, a_1, a_0$$

associated with U give place to the corresponding fundamental exponents

$$a_{N-1} + (N-1)h, \dots, a_1 + h, a_0$$

associated with V.

Consider the function Q_+ obtained from Q by adding on to the polynomials

$$Q_{N-1}, \dots, Z^{(N-2)h}Q_1, Z^{(N-1)h}Q_0\,,$$

as these latter appear in Q expressed as above in terms of V, arbitrary polynomials

$$q_{n-1}, \dots, q_1, q_0,$$

of degrees less than

$$0, \dots, (N-2)h, (N-1)\,h\,.$$

The special case of our theorem is now seen to apply to Q_+ expressed in terms of V. The number of conditions to apply to Q_+ to build it on (τ) at $Z = 0$ is, therefore,

$$N\{\lambda + (N-1)h\} - \Sigma a - \frac{N(N-1)}{2}h\,,$$

which reduces to the simpler expression

$$N\lambda - \Sigma a + \frac{N(N-1)}{2}h.$$

But, these conditions include the $\frac{N(N-1)}{2}h$ conditions involved in setting equal to 0 all the coefficients in the arbitrary polynomials (q). The number left over to apply to the arbitrary polynomials (Q) is, therefore, $N\lambda - \Sigma a$. In other words, the theorem holds in the general case.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLII : SERIES III : MAY, 1948
SECTION THREE

Illumination for Concealment of Ships at Night: Some Technical Considerations

By E. GODFREY BURR

Presented by A. NORMAN SHAW, F.R.S.C.

IN a previous paper (Burr, 1947), the development of this project was discussed and the extent of concealment attained in trial installations was noted. It is the purpose now to indicate the technical factors involved in this engineering development.

How nearly a practical installation could approach the ideal of a perfect brightness-match without any colour contrast, and consequently invisibility at any distance, can be estimated by consideration of the technical factors involved in approximating to these ideal conditions and the visual result of such approximation.

The ideal conditions are as follows:

A. Maintenance of the illumination of such areas at all times to give perfect brightness-balance with the background, which for the upper part of the silhouette is the sky, and for the lower part is sea, usually darker than the sky.

B. Uniformity of brightness over the areas of the silhouette artificially illuminated and absence of an appreciable colour contrast.

Failure to fulfil each and all of these theoretical requirements would result in detection of the pattern at finite ranges, which might be small or might approach the natural range according to the degree of deviation from the perfect theoretical condition.

In the following, a Foot-Lambert is the unit of brightness used.

PRELIMINARY APPROACH AND SHIPBOARD TRIAL INSTALLATIONS

In ordinary photometric measurements, very small deviations of the order of 1 per cent for either *A* or *B* are discernible, but the writer, from experiments in 1940, had formed the impression that, a very large reduction of the visual sense occurring under night conditions, much larger deviations could occur without rendering the pattern visible. This large reduction of visual sense had been made evident by the disappearance of a circular disc of maximum contrast of a visual size of one degree at a very low background brightness-level in the laboratory. This extreme case of an acuity of 1/60 of normal would, it was felt, have its counterpart in other criteria of vision such as the least perceptible brightness difference involved in *A* and *B*.

A. The technical considerations stated as *A* are those of control of the power of the lighting and development of apparatus for that purpose. The Optics Department at the National Research Council Laboratories co-operated in the early development of control apparatus.

B. The technical consideration stated above as *B* involves the engineering features of the illuminating installation, in its various aspects of electrical supply, light source, light distribution, including positive shielding of the light source under all conditions of rolling and pitching of the ship, the mechanical problem of support of overside units, and generally the problems of seaworthiness. Experimental shipboard installations would be the most direct approach.

This subdivision of the work was generally followed throughout. Work was undertaken for *B* by an experimental installation of the corvette H.M.C.S. *Chambly* in March, 1941, which also provided for study of colour contrast by provision of blue, blue-green, and clear filters for the experimental luminaires employing standard 25-watt "Mazda" lamps. Manual rheostatic control operated by signal from the observing sea-boat was relied on for *A*. The light distribution control was by holophane prismatic plates giving a fan-shaped beam; metal cowls were used to cut off the light completely at angles below the horizontal seaward, 20° fore and aft, and 40° abeam, to allow for pitch and roll of the ship.

TECHNICAL CONCLUSIONS FROM THIS FIRST SHIPBOARD INSTALLATION

1. *Mechanical.* It was decided that seaworthy supports for luminaires could be readily made. Those for the overside units would be retractable, to allow of stowing inboard when desired, which feature had been carried out in this first installation. The weight of the supports would be reasonable but could be reduced if a smaller and lighter luminaire were produced.

This opinion was accepted by the sea-going officers closely associated with this installation, but did not receive general acceptance until August, 1943, when it was found that the supports of a later installation incorporating these features were seaworthy in operational duty and that the luminaires, which developed some defects, could readily be improved.

2. *Colour contrast.* In these experiments it was found that on moonless nights the colour filters did not appreciably affect the results, but that with full moon the ship appeared pink-tinted just inside the visibility range without the use of a filter. The blue filter was too dense to allow of a balance in full moonlight, but the blue-green gave a better effect than the clear glass under all the night conditions tested. In later installations, a blue-green filter was used, which was too green. A daylight blue is believed to be most suitable.

3. *Extent of coverage by illumination of the ship silhouette.* It was concluded that the conditions of installation on a warship make impossible the illumination of all of the silhouette.

The many small deck structures, ammunition lockers, gun shields, depth charges, etc., present surfaces which are extremely difficult to illuminate uniformly and the necessary zones of fire of guns prevent the installation of supports for luminaires.

It was soon realized that, if these consequently dark areas were of small size relative to the whole and were surrounded by lighted areas, they merged into their surrounding at distances well within the range of detection of the whole ship. As these formed small areas of maximum contrast, the merging effect was obviously due to their being within the critical acuity angle which was known, for maximum contrast, to be of the order of four minutes for moonlight and sixty minutes under extremely dark conditions. The range at which they would be separately visible would be in the direct ratio of their significant dimension to that of the whole ship.

Many of these dark areas were surrounded by lighted areas, but at the upper edge of the silhouette it was necessary to leave an irregular band of unlighted area, and also, at some aspects, dark areas normally well separated would be brought together.

4. *Detection of the dark residue of silhouette chief factor of practical design.* It was observed that, frequently, detection resulted from seeing the dark residue of the silhouette only; on other occasions a dim shadow of the remainder was also visible and occasionally a slight glow or a dip in the horizon was detected. It was concluded that the visibility of the aggregate of the unlighted residue areas is the chief factor limiting the amount of range reduction practically obtainable, but small detached unlighted areas merge and then affect visibility only if the mean brightness of the merged area (within the acuity angle) is too high or too low for proper balance against the background.

5. *Significant dimension for acuity of irregularly shaped areas.* At that time it was not known whether the minimum, the mean dimension, or the square root of the area was the significant dimension. Later, reliable information was obtained that all warship silhouettes at all aspects gave, other things being equal, ranges directly proportional to the square root of their area within a ± 5 per cent variation, although a ratio of major to minor axes of more than 10 may exist. From these sea experiments it was expected, and later it was found by experiment to scale in the laboratory on the actual patterns of the residual dark upper band, that the above relation does not hold for these patterns which are of irregular width. In previous experiments on aircraft silhouettes, it had been noted that, at the threshold of detection, only the stubs of the narrow projecting wing areas were visible, but it was not known whether this effect did in fact, by reducing the visible area, reduce the detection range of the pattern.

Later, in the laboratory, the ship residual silhouette also showed similar visual effects; the narrower parts of the band disappear and leave at the threshold of detection one or more dark areas. This visual change in the pattern eliminates a considerable part of the area completely at threshold and divides the remainder into separate areas each of which has its own threshold range, corresponding to but a small part of the total area.

In a specific case, the residual band included one-quarter of the silhouette area and should have been visible at one-half the natural range. The actual range was one-quarter and two separate areas were visible near threshold range. Eliminating either had no appreciable effect on range.

6. *Significant dimension of dark residual area.* It appears that, with such elongated areas of irregular width, the threshold range is more nearly related to the square root of the largest compact area than to any other dimension. The natural ship silhouette while irregular in outline has its main compact area centrally situated and, while excrescences may disappear, the effect of area reduction at threshold is within the experimental error. These residue patterns have, however, more than one compact area joined by considerable length of relatively narrow areas and appear split up, at the visual threshold. Aspects fine on the bow give residual areas which have a single dominating compact area but beam aspects give more than one compact area.

The practical limit of coverage by the lighting appeared to be about 75 per cent of the total area which on the straight area basis is a range reduction ratio of 2, but the largest compact area could be reduced

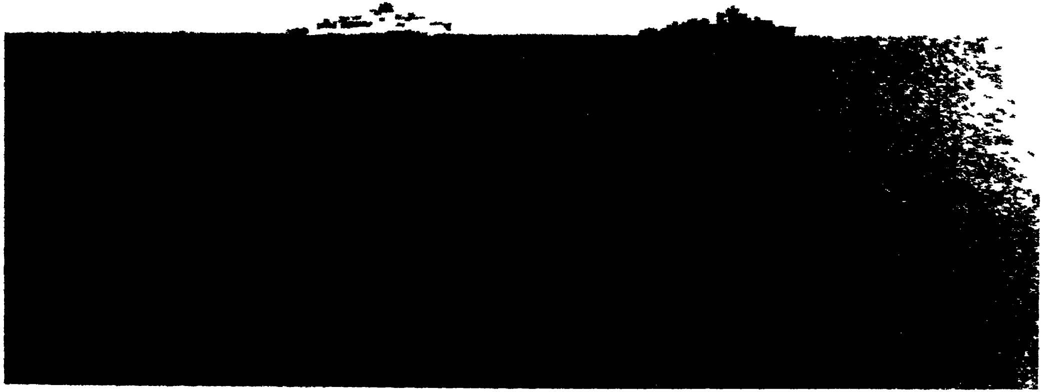


Figure 1 —Natural and illuminated silhouettes of a frigate aspect 45° from bow

well below one-quarter of the whole. Two compact areas of one-sixteenth the total for the beam and 45° aspects is possible in many cases giving a range reduction ratio of 4 to 1.

Few occasions would arise in service where a greater reduction would be useful, particularly as hazy conditions reduce range and range ratio greatly.

7. *Balanced design.* A desirable design would be obtained if the chances at the threshold were equal for detecting either of the dark residual compact areas, or the illuminated area as a light patch; at one-quarter of the natural range with 100 per cent atmospheric transmission. Fig. 1 shows the appearance of such a ship compared with the ordinary ship and, viewed in subdued light, a difference in the detection ranges of the two ships will be evident. The degree of control of balance of the lighting to equalize these chances must then be determined.

A. TOLERANCE OF RESIDUAL CONTRAST

1. *Tolerance of residual contrast of the illuminated area.* The above requires that the illuminated area shall not have a visible contrast either dark or bright against the background when viewed at one-quarter of the natural range.

There was but little data available as a guide to the allowable residual contrast at threshold in night vision. Most workers in this field had used the method of constant distance and variable brightness and pattern size in establishing threshold acuity and while noting the fact that contrast and pattern size were related, such data were presented in the form of minimum-observable contrast as a function of background-brightness for a given pattern size. The threshold pattern size (angular) as a function of the pattern contrast at constant background-brightness if available would be directly applicable in terms of range and contrast and hence determine the allowable tolerance of control for a perfect brightness-match.

2. *Visibility of dark and bright patches.* In the middle of 1942, the General Electric Company collected available published data of threshold angular size from tests on simple compact geometrical patterns of varying contrast (dark) against a large outer field of constant brightness at night vision levels. It has been assumed that the data apply to circular dark patches.

At the same time the National Physical Laboratory data on threshold ranges of ships at night as a function of contrast of silhouette against sky were made available by the Admiralty (Smith, 1942).

Tests of the visibility of circular light patches at two levels of night sky brightness made by the National Physical Laboratories were also made available (Meetham and Lambert, 1941).

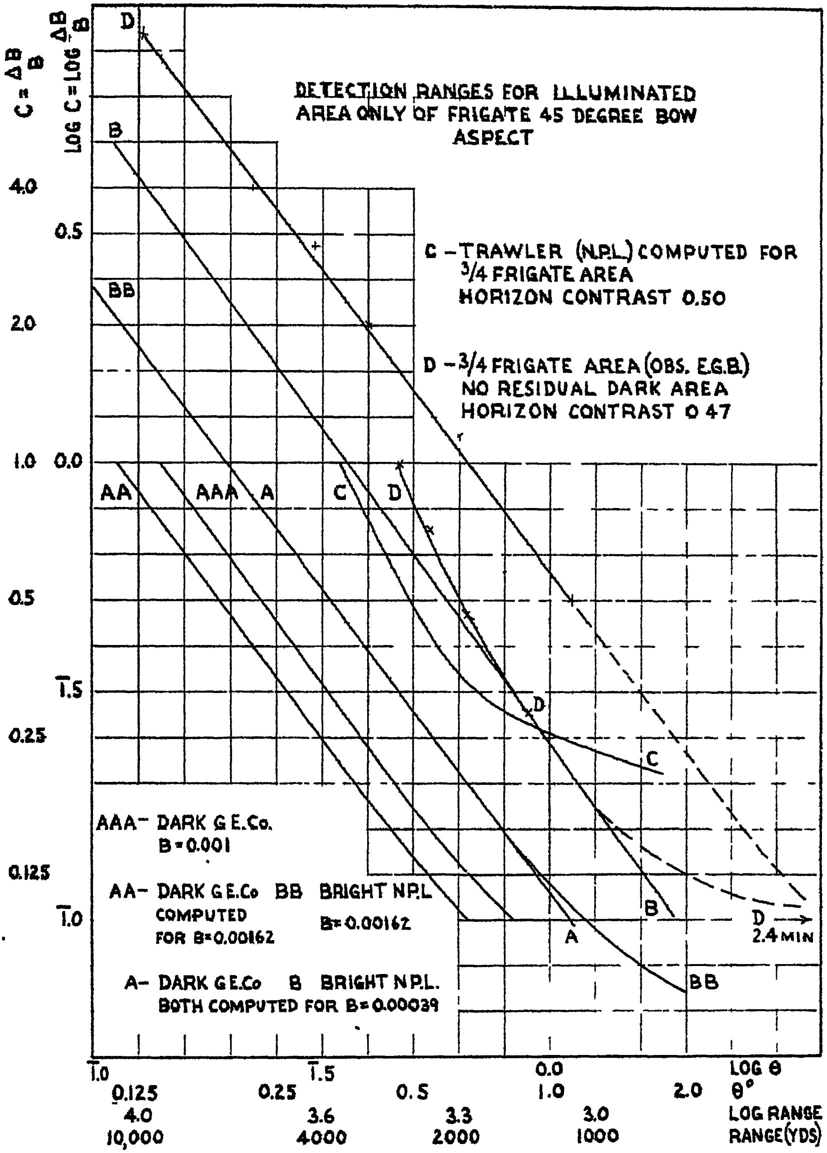


Figure 2.—Acuity data for dark and bright contrast.

It should be noted that the National Physical Laboratory ship results are for the complex pattern of a trawler silhouette seen at the junction of a background of sky of the stated brightness and sea of one-half that brightness (50 per cent contrast)

The General Electric Company dark patch and National Physical Laboratory light patch results are for simple compact patterns seen against a uniform background. In all three cases 100 per cent atmospheric transmission is assumed. These three sets of data are plotted as Fig. 2. The original data are replotted for the common background-brightness value of 0.00039. The co-ordinates are visual angle and range of three-quarters of the silhouette area of a frigate (the illuminated area) as a function of the dark or bright contrast $\pm\frac{\Delta B}{B}$ to logarithmic scales. Curve *BB* is the original data (light patch) and curve *AA* is the General Electric Company dark patch computed for the same background-brightness of 0.00162. It should be noted that within these limits of angular size each is nearly a straight line of the same slope with a constant difference of log *R* and log θ of 0.24. The inference is that at this brightness-level, other things being equal, a light patch is less visible than a dark patch in the range ratio of 57.5 per cent.

It had been assumed that equal small increments and decrements of brightness would be equally visible.

The curves compared are for original National Physical Laboratory data of a circular bright patch at $B = 0.00162$ and the General Electric data for $B = 0.001$ corrected to $B = 0.00162$.

The curve of Fig. 3 was used for this purpose and is in close agreement of shape with the National Physical Laboratory ship pattern curve.

This surprising difference in visibility of circular patches of the same dark and bright contrast needed further confirmation

An experiment with a frigate pattern (scale 1/64 in. equals 1 ft.) was undertaken. The 75 per cent illuminated area was made transparent and its brightness was made separately controllable from zero to about ten times the sky brightness ($B = 0.00039$) against which it was observed at a 50 per cent contrast horizon.

The observer was the writer, whose night visual acuity is about 25 per cent below normal.

Curve *D* (Fig. 2) shows the results with all observations indicated from 100 per cent contrast dark to about nine times bright. A difference between dark and bright visibility is evidently of the same order of magnitude as above.

The difference is not constant since the dark contrast curve is not a straight line but follows rather the form of the National Physical

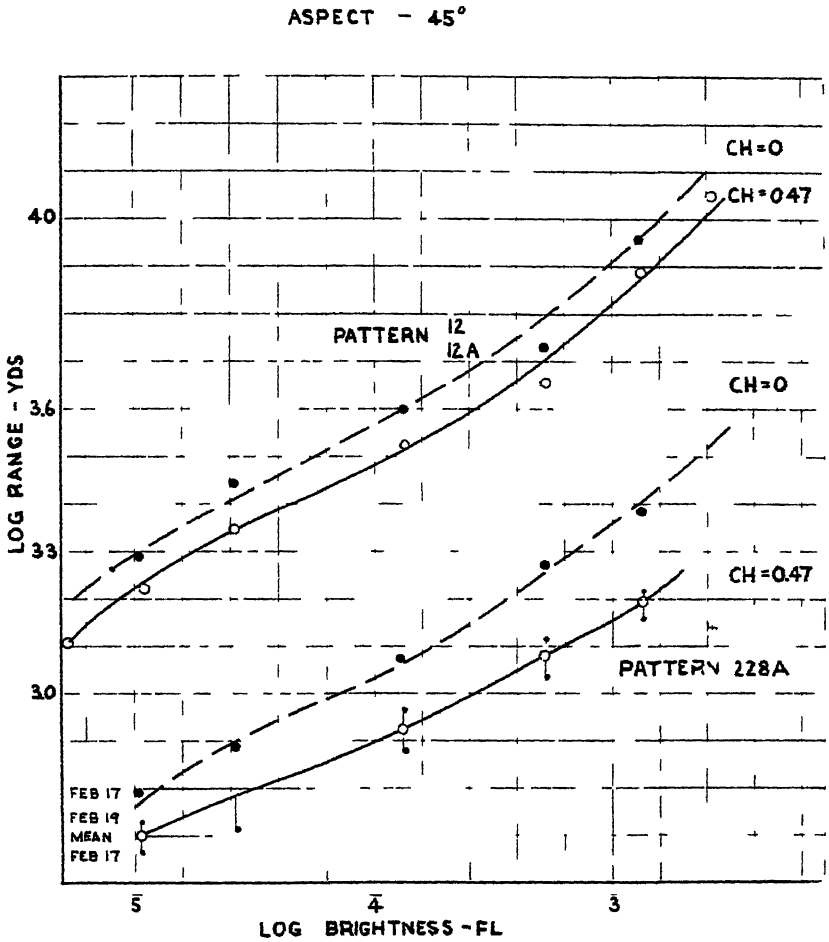


FIGURE 3 —Comparison of optimum illuminated and natural patterns at various brightness-levels.

Laboratory ship pattern curve *C*, but with relatively greater visibility at lower contrast. This deviation may be due to the horizon contrast of the background against which the ship patterns are seen.

From the curves of Fig. 2, the following provisional values of contrast tolerance for 4/1 range reduction from a natural contrast of 80 per cent may be obtained.

For visibility of the illuminated area only
(75 per cent of total)

Curve *A* – G. E. Co. data (dark)		Log $R = 3.73$	log $C = \bar{1}.9$	$C = 0.8$
		Log $R = 3.13$	log $C = \bar{1}.1$	$C = 0.125$
Curve *B* – N.P.I. data (bright)		Log $R = 3.13$	log $C = \bar{1}.44$	$C = 0.25$
Curve *D*	dark	Log $R = 3.4$	log $C = \bar{1}.9$	$C = 0.8$
		Log $R = 2.8$	log $C = \bar{1}.1$	$C = 0.125$
	bright	Log $R = 2.8$	log $C = \bar{1}.3+$	$C = 0.20+$

The tolerance of control should not exceed 12.5 per cent dark and 25 per cent bright if detection of the illuminated area only were the criterion.

3. *Inclusion of residual dark areas in the estimate.* Fig. 4 shows the result of adding the residual dark areas in an experiment similar to the one described above as curve *D*, Fig. 2.

Curves *A*, *B*, and *D* are the same as in Fig. 2.

Curve *E* is with the residual dark area added (contrast 0.8). It will be noted that the range is increased for dark contrast but is reduced for bright contrast of the illuminated area against the sky.

The minimum range is for a ship brightness of 1.35 times the sky brightness. This is an average brightness ratio of

$$\frac{(1.35 \times 0.75) + (0.2 \times 0.25)}{1} = \frac{1.06}{1}$$

or nearly zero average contrast.

4. *Average contrast of the whole ship pattern.* The upper graph is a replot of curve *E* (two trials) for contrast measured from the balance points at 1.35 × *B* sky, with added points from a sea trial on the corvette H.M.C.S. *Rimouski* observed abeam. Any difference in area of its silhouette and that of the frigate at 45° bow aspect used in the laboratory experiment, is compensated by the difference of acuity of the writer.

Fair agreement is to be noted except for the low dark contrast where portable photometric observations would be subject to large error.

It is clear that some overlighting can be usefully employed within uncertain limits to compensate for the dark residue to an average balance against sky.

In Fig. 5, all the above experiments have been replotted with average contrast for the whole ship area using logarithmic scales with the range scales as ordinates. For the bright contrasts all the laboratory curves follow closely the same law as the National Physical Laboratory curve with a nearly constant difference in log R of -0.2.

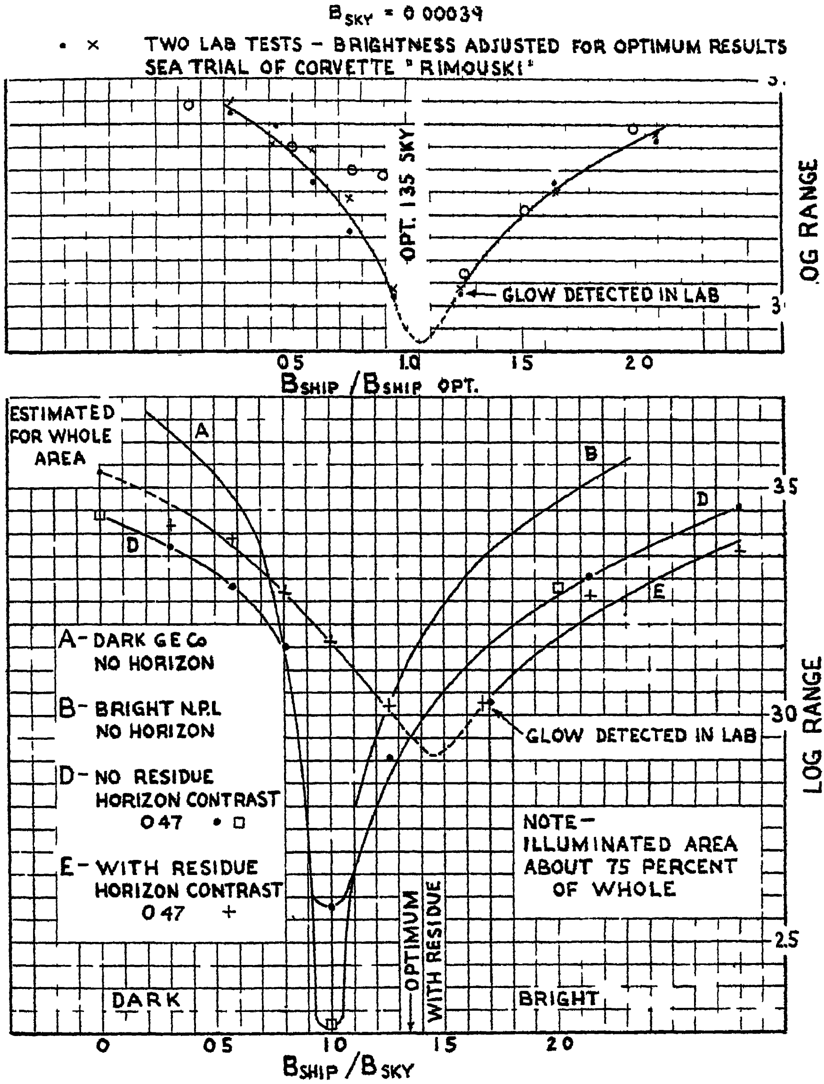


FIGURE 4.—Overlighting as compensation for residual dark areas.

The reduced acuity of the observer (E.G.B.) is known to amount to 0.1. The three corvette sea trial values agree even more closely with the National Physical Laboratory curve. For these, the observers

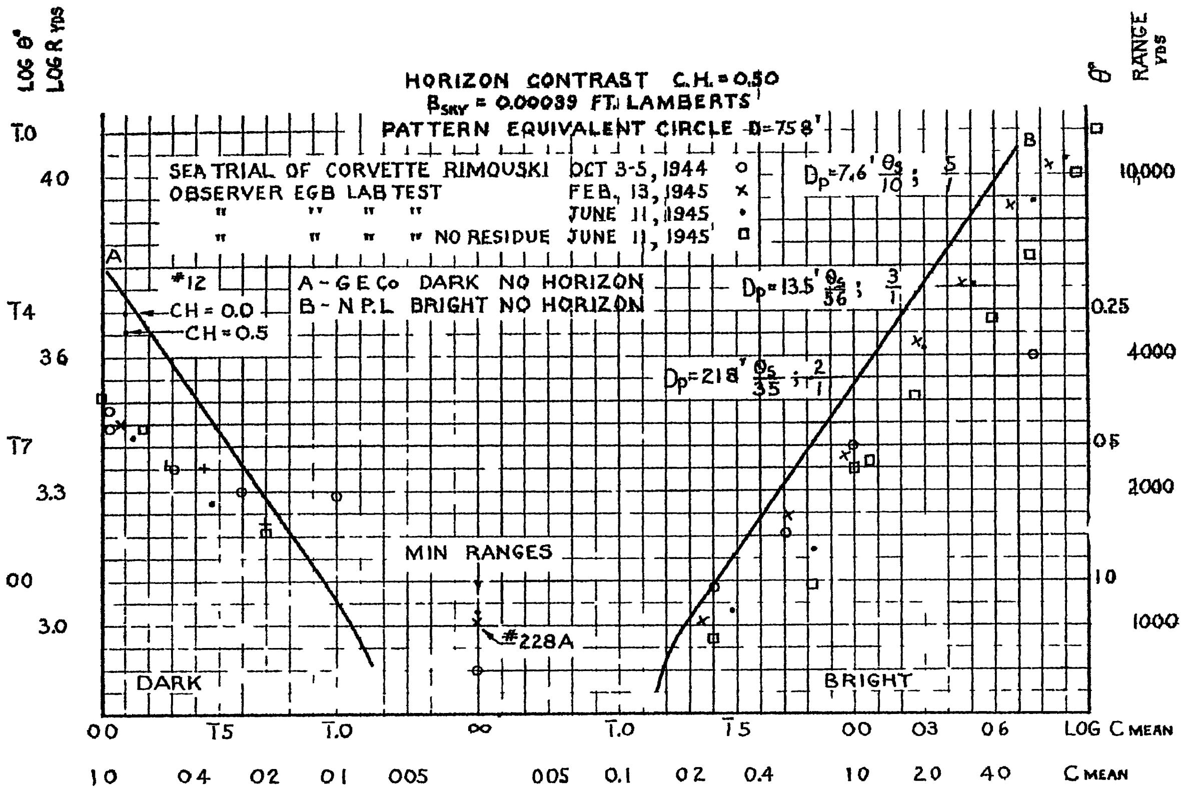


FIGURE 5.—Range of visibility of frigate versus mean contrast of whole ship.

were three watch-keeping officers and agreement is within the probable error. However, there appears to be a slight systematic difference from the National Physical Laboratory curve giving slightly reduced ranges as no ranges greater than the National Physical Laboratory data were obtained.

The National Physical Laboratory data may then be used for a conservative estimate of range due to excess brightness without consideration of horizon contrast.

For dark contrast, these laboratory tests, and also the sea trial, give ranges well below the General Electric Company curve at the higher contrasts but approach at the lower contrasts A similar effect is shown by the National Physical Laboratory trawler curve, Fig. 2. This shortening of range is greater than the reduced acuity of the observer. In other laboratory tests using a better than average observer, the frigate pattern (45° aspect) 12*A* gives almost perfect agreement with the General Electric Company curve for 80 per cent dark contrast and a horizon contrast of zero (see Fig. 5). Also the difference of range for a change of horizon contrast from zero to 0.5 is small for this high contrast pattern as shown on Fig. 5.

The reason for these discrepancies is not clear but as all ranges are below the General Electric Company curve this curve may be used for a conservative estimate of range due to dark contrast.

5. *Average contrast tolerance for stated brightness.* For the above-stated case of range reduction ratio of 4 and sky brightness 0.00039 the tolerance of control is 10 per cent dark and 25 per cent bright, average contrast of the whole ship from the optimum balance point.

6. *Average contrast tolerance for full moonlight.* The above difference in tolerance, dark and bright, is based on data at $B = 0.0016$. Only the General Electric Company data (dark) extends to the full moonlight levels. At $B = 0.01$ the curve has the same slope and so gives the same tolerance of 10 per cent for 4/1 range reduction. As there is evidence that at this level the least perceptible increase and decrease of brightness may be equal, to be on the safe side a bright tolerance of 10 per cent is chosen. The set-up did not allow of experiment to verify this assumption.

7. *Accepted estimate of average contrast tolerance for control equipment.* The control mechanism should maintain an average brightness-match within a tolerance of ± 10 per cent when operating from moonlight down to starlight levels. For very dark night levels the tolerance is greater, but must not exceed ± 20 per cent.

8. *General conclusion regarding visibility.* From the above, general conclusions can be drawn as follows. The range of visibility of an

illuminated ship is related to the average contrast of the whole pattern following established relations except that at a certain minimum range the residual dark area or areas become separately visible even if the mean contrast be reduced to zero. This will occur at the threshold range of the largest compact dark area determined from established data.

B. ALLOWABLE DEVIATION OF BRIGHTNESS FROM THE AVERAGE WITHIN THE ILLUMINATED AREA

The previous paragraph establishes residual average contrast tolerances of 10 per cent dark and 10 per cent bright for a four-to-one range reduction of the whole pattern on the assumption that the illuminated area is uniformly bright and the residual area is uniformly dark with 80 per cent contrast against the sky. As the General Electric Company and National Physical Laboratory curves are threshold curves giving the angular size of a dark or bright patch of stated contrast, they may be used to determine the threshold size of such patch as a fraction of the whole silhouette area when its brightness difference and hence contrast is stated.

Considering the possibility of detecting separately the areas of maximum brightness of the illuminated area the National Physical Laboratory curve gives the following result (Fig. 5).

At the reduced range (4/1) the whole ship pattern subtends an angle of 1.25° and its threshold of visibility is an average bright contrast of 19 per cent.

Similarly, an area subtending an angle of 0.36° is at the threshold of detection if its bright contrast is 100 per cent, that is to say if its brightness is twice the average. The ratio of the above angles is the ratio of the diameters of the equivalent circles on the actual silhouette, namely 3.5.

1. *Ratio of maximum to average brightness and diameter of patch.* It is therefore possible to specify the allowable variation of brightness of the actual illumination scheme in terms of ratio of excess to mean brightness and the diameter of the circular area of excess brightness.

In applying such a method to the practical case a safety factor of two is introduced as it is desirable that such detection of bright spots should not occur at the detection range of the camouflaged ship as points of light are conspicuous at sea. Lacking such a conspicuous appearance it is highly probable that nearer approach might be possible than the staged trials would indicate as at threshold the appearance is different from that of an ordinary ship but not conspicuous.

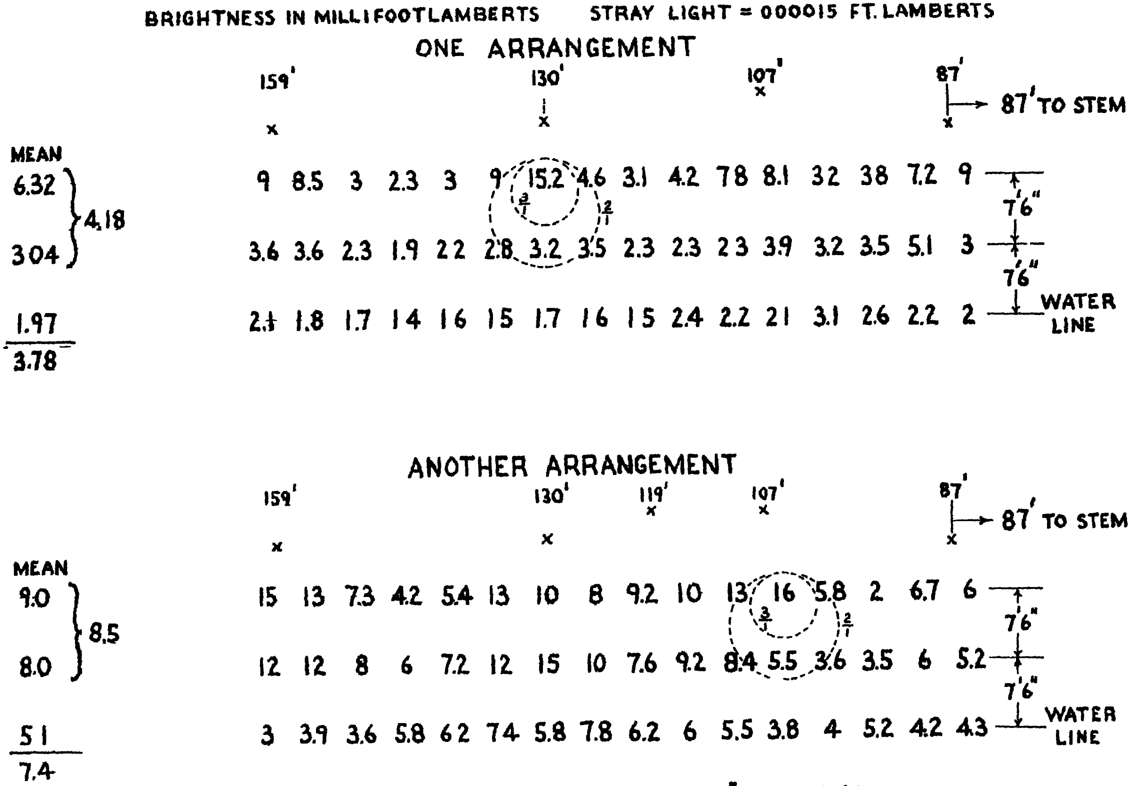


FIGURE 6.—Part of frigate hull brightness-survey, starboard side.

The factor of two would allow of closing the range to one-eighth of the natural range before detection of bright spots would occur.

The following values would then apply:

$\frac{B \text{ excess}}{B \text{ average}}$ =	2	3
Diameter of patch (ft.) =	11	6.75

2. *Application to brightness-survey of ship.* Fig. 6 shows brightness-surveys of two arrangements of luminaires for a 75-foot part of a frigate hull with the above circular patch areas shown for the locations of maximum brightness. It will be seen that this practical layout gives, for the points of maximum brightness, patch areas smaller than the corresponding patch size for detection.

This area would therefore show no points of light at one-half the detection range.

Sufficient allowance for improved bright contrast perception in full moonlight seems to be provided by the above factor of two, and the use of the National Physical Laboratory curve which appears to represent maximum perception of bright contrast.

Similar consideration of dark patches of great contrast in the illuminated area would give a patch size of one-tenth the equivalent circle of the whole silhouette or 8-foot diameter and corresponds to a brightness of the patch reduced one-tenth of the mean brightness (contrast 0.9). The survey shows no areas of this size with brightness so low, the lowest being about one-quarter of the mean brightness. Again we can predict that at one-eighth of the natural range, no dark patches will be visible in the illuminated area. Lacking any visible light or dark patches it will appear uniformly bright at all greater ranges.

It is true that these surveys were made on the part of the hull where small plate curvatures are favourable to small variation of brightness, but even in the case of more complicated surface shape the areas involved in large contrast either dark or bright are small relative to the critical values shown by these curves.

It may then be said that the type of luminaires developed are capable of producing a brightness-distribution which will appear uniform at one-eighth the natural range, at which range the unlighted residue and the illuminated area will be separately visible.

3. *Use of night binoculars.* The use of night binoculars increases the natural range of dark patches but by less than the magnification ratio (Smith, 1942).

The visibility range of light patches may more nearly increase with the magnification ratio but factual data are lacking.

It is therefore possible that, with night glasses, detection of bright patches may occur at somewhat greater relative ranges than as discussed above for naked eye conditions. Such an effect would reduce the distance between the detection range and the range at which the illuminated area is resolved into areas of differing brightness, but it is believed that some considerable difference would still exist sufficient to avoid detection of patches and a conspicuous appearance near the detection range.

SUMMARY

A. A control equipment capable of maintaining an average brightness balance within a tolerance of ± 10 per cent is required for a range reduction ratio of 4.

B. Adjustment of brightness to a sufficient uniformity can be made by applying the method of area and brightness ratio test.

ACKNOWLEDGMENTS

Many persons, naval and civil, helped this project. Lt.-Com. R. McCallum, R.C.N.V.R., was associated throughout with all sea installations and trials. He accompanied H.M.C.S. *Edmundston* in operational duty. Lt. D. Armstrong, R.C.N.V.R., did a like service on H.M.C.S. *Rimouski*. In 1944 they, with W. J. N. Vanderleck and other members of the Electrical Engineering and Optics Departments of the National Research Council, designed and built a gyro-stabilized photoelectric trainable automatic controller to conform to tentative staff requirements stated by the Admiralty in August, 1943. Sub.-Lt. H. Howell, R.C.N.V.R., was for a few months an invaluable observer in the laboratory. Sub.-Lt. McLellan, R.C.N.V.R., constructed the full-size laboratory sections for adjustment of the illumination and made the brightness-surveys. Messrs. Halvorson and Breeding of the General Electric Company were responsible for the development work of the luminaires and control done by that company, and collaborated in our service installations. Capt. J. D. Prentice, D.S.O., D.S.C., R.C.N., was at all times a friendly adviser and a staunch supporter.

REFERENCES

BURR, E. GODFREY (1947). Illumination for concealment of ships at night. Trans. Roy. Soc. Can. **41**: 45-54.

MEETHAM, A. R. and LAMBERT, G. E. V. (1941). Visibility of light patches at low illumination. London, National Physical Laboratory.

SMITH, T. (1942). An experimental study of the utility at night of binoculars of different powers, etc. London, National Physical Laboratory.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLII : SERIES III : MAY, 1948
SECTION THREE

The Determination of the International Volt to within Twenty Microvolts by the Application of Ageing Corrections to Old Normal Weston Standard Cells

By A. HENRIKSON and A. NORMAN SHAW, F.R.S.C.

INTRODUCTION

In an investigation by A. N. Shaw, H. E. Reilley, and R. J. Clark (1930) the ageing of Weston normal standard cells was shown to follow very closely the relation,

$$E_M = A + B \log (M + \tau) ,$$

where E_M is the difference from an initial reference mean of voltage, A, B, and τ are constants for the particular batch of cells being examined, and M is the age of the reference group. It was then found by application of the ageing correction determined by this equation, that the international reference mean for voltage could be recaptured for a period up to twenty years to within a few microvolts. As the results given below constitute an interesting sequel to work first described to the Royal Society of Canada many years ago, it was thought appropriate to report them in these transactions.

The objects of this investigation were: (1) to test the condition of the collection of old cells which had been previously compared with standards at London and at Washington on various occasions since 1908; (2) to check the ageing correction for Weston normal cells over a period of forty years; and (3) to determine the accuracy with which the international reference mean for voltage could be recaptured throughout a long period by a small collection of cells, and without necessarily making new cells or new comparisons with other standard laboratories.

The apparatus used is essentially that set up by one of the authors in 1909. It consists of the usual equipment for this type of work, a thermostatically controlled oil bath in which the cells are immersed, the temperature being kept constant at 25.00° C. to within 0.01 degree, and a simple slide wire potentiometer along which a potential drop of one microvolt per millimetre is steadily maintained. An old Thomson type galvanometer which gave a sensitivity of about 1 microvolt per scale division provided the null indication.

The cells are opposed to each other in pairs, and the successive differences in electromotive force between cells are measured, one cell being chosen as a reference. The method of treating these readings in order to recapture the initial mean is given fully in a paper by Shaw and Reilley (1930).

EXPERIMENTAL OBSERVATIONS

Recapture of an Initial Reference Mean with Cells over Forty Years Old

In order to recapture an initial reference mean, a record of the electromotive force of the cells at an early period of their existence must be available. From these records the constants in the ageing equation may be determined and the correction obtained for any subsequent date. The routine for doing this is given fully in the above-quoted paper.

Cells C2, 1.6, and 1.8 are our only cells to survive in both reliable and satisfactory condition for a period of nearly forty years, and an attempt was made to predict their present mean value in terms of the initial reference mean, that is, our 1911 estimate of the international mean for voltage. These cells were made in September, 1909, but adequate records of their values before 1912 are not available. Thus it is only possible to obtain an indirect estimate of the value of τ in the ageing equation. An ageing equation can be obtained, however, from the data of Table I, which gives the records of these cells on three widely separated dates.

TABLE I

AVERAGE EMF OF CELLS C2, 1.6, AND 1.8 REFERRED TO THE INITIAL MEAN

Date	July, 1912	May, 1919	Aug., 1925	Dec., 1947
Age	34 months	116 months	191 months	predicted value 459 months
	−13	−19	−26	−47

If τ is assumed to have the value 0.75, substitution of the above values in the ageing equation ($E_M = A + B. \log (M + \tau)$), gives an average value for A of 57.8. Thus the ageing equation for these cells is: $E_M = 57.8 - 39.4 \log_{10}(M + 0.75)$. This gives an ageing correction for December, 1947, of −47 microvolts. This is the most probable value of the correction for these cells, since τ is usually found to be less than 1.

If τ had a value as high as 25, the highest value found in the past for any of our cells, the ageing equation would become: $E_M = 62.9 - 39.4 \log_{10}(M + 25)$, giving a correction of −43 microvolts. Thus a maximum error of only 4 microvolts could be made by using an average value for τ.

The present value of these cells, as determined from the ageing equation was, in this way, taken as −47 microvolts referred to the initial reference mean. Or, in other words, the initial reference mean was estimated to be 47 microvolts above the present electromotive force of these cells. Our "initial reference mean," was the mean value of a particular group of cells in 1911 which at that time corresponded very closely to the international mean of voltage.

In order to test the ageing equation for the old cells, several methods may be employed: (*a*) comparisons with younger neutral cells, (*b*) comparisons with acid cells, and (*c*) direct comparisons with other laboratories.

Recapture of Initial Reference Mean Using Younger Neutral Cells, and Acid Cells

Table II gives the mean of a number of neutral cells constructed in 1923. These values were checked by several direct international comparisons, and are therefore also expressed in terms of the 1911 international mean of voltage.

TABLE II

RECORD OF THE AGEING OF THE REFERENCE CELLS (1923 BATCH)
MEAN OF CELLS 60, 62, 64, 65, 66, 67, 69, 71, AND 79

Age	0.5 months	2 months	8 months	20 months	40 months	60 months
Mean	+38	+20	+4	−12	−21	−30

In obtaining the average for each date, one of the "selection rules" (Shaw, Reilley, and Clark, 1930) has been applied; namely, reject any cell whose deviation from the mean electromotive force of the batch is greater than $20 + d$, where d is the mean deviation.

All neutral cells made in 1923 were found to have an average value for τ of 0.75, and substitution of this value, and the use of the data of Table II in the ageing equation give:

$$E_M = 41.5 - 39.4 \log_{10}(M + 0.75) .$$

This gives an average correction for these cells at the present time, age 290 months, of −56 microvolts.

The present mean of the old 1909 cells is 3 microvolts higher than the mean of the 1923 batch and is therefore −53 microvolts, referred to the international mean of voltage. This compares favourably with the value −47 microvolts obtained by consideration of the ageing of the old cells alone.

It was shown (Shaw, Reilley, and Clark, 1930) that acid cells remain more nearly constant for many years, with only a very small decrease of electromotive force, and from our former McGill results and calculations this would become about −71 microvolts in twenty-five years, that is, at the present time. By comparing two steady acid cells with the 1923 neutral cells to which the ageing correction had been applied, the value of the acid cells is found to be the same as the predicted value, −71 microvolts.

Table III gives a summary of the above comparisons. The last two columns in this table are the value of the old cells in terms of the international mean of voltage as determined through comparisons with, respectively, younger neutral cells and the acid cells, duly corrected for ageing.

TABLE III

SUMMARY OF CELL COMPARISONS

	In terms of international mean as determined by ageing equation	Value as obtained by corrected comparisons with the 1923 neutral cells	Value as obtained by corrected comparisons with acid cells
The old 1909 cells	−47	−53	−53

Recent Cell Comparisons with the Bureau of Standards in Washington, in January, 1948

In order to test the results of applying the ageing equation to the old cells discussed in a previous section, and also the approximate methods based on comparison with the means of both acid and neutral cells, also respectively corrected for ageing, six cells were carried to the National Bureau of Standards in Washington, D.C., in January, 1948. These cells were placed in a temperature regulated oil bath, kept at 20.00° C. for several days and measured with a Leeds and Northrup Standard Cell Comparator. The results of these measurements are given in Table IV. Also included in this table are the values of these cells as measured at McGill at 25.00° C., before and after having been taken to Washington. The difference between the McGill and Washington measurements is also given. The change from 25.00° C. to 20.00° C. does not appreciably affect the differences among the cells at any one temperature in this range.

TABLE IV

SUMMARY OF INTERNATIONAL COMPARISONS

Cell	Referred to McGill mean before trip	Referred to McGill mean after trip	Referred to Bureau in Washington	Difference between Washington and McGill means	
60	−81	−77	—	—	
62	−62	−47	−18	−29	
63	−58	−56	−25	−31	−27
65	−61	−51	−31	−20	
66	−70	−57	−44	−13	
69	−43	−45	−36	− 9	−11
83	−65	−68	−58	−10	

Since a slight shaking of the cells occurred accidentally during the journey to Washington, it is believed that only the readings taken at McGill after the trip should be compared with the Washington values. If the mean of all the cells is used, the McGill reference mean is found to be 19 microvolts higher than the United States reference mean. However, since the values of some of the cells appear to have changed erratically, relative to the others, probably due to the shaking in transit, it is believed that only those that have remained nearly steady relative to one another should be considered. On this basis only the last three cells should be used, giving *an average difference of only 11 microvolts between McGill and the Washington reference means*. It should be noted that the McGill reference mean, cited here, is that determined by the 1923 batches of neutral and acid cells. Based on the old 1909 cells, the McGill reference mean is 6 to 10 microvolts lower. Table V will serve to clarify the above statements and to indicate the relative positions of the initial reference mean, the Bureau of Standards reference mean, and the present value of the 1909 cells in terms of these means.

TABLE V

SUMMARY OF RESULTS IN TERMS OF THE INITIAL REFERENCE MEAN CHOSEN IN 1911

Difference in microvolts	
0	Present McGill reference mean obtained by analysis and ageing corrections from the records of the 1923 neutral and acid cells, and also assumed to differ negligibly from that in use at the National Physical Laboratory in Teddington, England.
−6	Present McGill reference mean obtained by application of the ageing correction to 40-year-old neutral cells.
−11	Present Bureau of Standards reference mean, obtained by recent comparisons.
−19	Present Bureau of Standards reference mean, alternative estimate.
−53	Present mean actual value of the 40-year-old neutral cells, as determined directly with the aid of the 1923 neutral and acid cells.

It is therefore seen that the use of these old cells to determine the international mean of voltage would, after applying ageing corrections, produce an error of only about −6 microvolts, as compared with the estimate of the initial mean based on the 1923 neutral and acid cells, and an error of at least +5 microvolts (up to perhaps +13) as compared with the mean at present used by the Bureau of Standards.

The old cells had a high internal resistance which made it very difficult to obtain a close balance on the potentiometer; there may be a possible error of ±5 microvolts in any single observation. However, if the estimates were based only on the three very old cells, it seemed still possible to recapture our estimate of the international mean of voltage to within 10 microvolts and that of the Bureau of Standards to within 15 to 20 microvolts. If a larger number of old cells were kept available with complete records of their early behaviour, the international mean could probably be recaptured at will, with these methods, to within only a few microvolts, over at least a period of thirty years.

Dr. G. W. Vinal, and Dr. W. J. Hamer, at the Bureau of Standards, gave valuable advice, facilities in their cell laboratory, and a cordial welcome in Washington. It is a great pleasure to acknowledge their kindness and co-operation.

McGILL UNIVERSITY

REFERENCES

SHAW, A. N. and **REILLEY, H. E.** (1930). Can. J. Res. 3: 473-89.

SHAW, A. N., **REILLEY, H. E.**, and **CLARK, R. J.** (1930). Phil. Trans. Roy. Soc. London, Series A, 229: 125-62.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLII : SERIES III . MAY, 1948
SECTION THREE

On Perron's Extension of the Euler-Knopp Summation Method

By M. S. MACPHAIL

Presented by R. L. JEFFERY, F.R.S.C.

THE Euler-Knopp series transformation carries Σa_n into ΣA_n, where A_n is defined by

$$\mathfrak{E}(r): \qquad A_n = r \sum_{k=0}^{n} \binom{n}{k} r^k (1-r)^{n-k} a_k,$$

provided $r \neq 0, 1$; $\mathfrak{E}(1)$ being defined as the identical transformation and $\mathfrak{E}(0)$ remaining undefined. If $\Sigma A_n = S$, we say that Σa_n is summable $\mathfrak{E}(r)$ to S. (All summations run from 0 to ∞ unless otherwise stated.) This summation method is regular if and only if $0 < r \leqq 1$, but it is consistent with convergence for all $r \neq 0$, by which is meant that if Σa_n and ΣA_n both converge, they converge to the same value. The main properties were established by Knopp in 1922-3 and Agnew in 1944, the former treating only the case $0 < r \leqq 1$, while the latter allowed r to take any value. It was pointed out by Perron in 1923 that $\mathfrak{E}(r)$ can be generated by the function

$$F(t) = \frac{rt}{1-(1-r)t}$$

in the following way. Write down formally the series

$$\sum_{n=0}^{\infty} a_n [F(t)]^{n+1},$$

and arrange it in powers of t. It is easily verified that this gives $\Sigma A_n t^{n+1}$, from which the $\mathfrak{E}(r)$ transform ΣA_n is obtained by putting $t = 1$. If in this we take t as a complex variable, we can use standard methods of analysis to obtain short proofs of many properties of $\mathfrak{E}(r)$. We have also an obvious generalization, namely, to replace $F(t)$ by any function

$$f(t) = \Sigma c_n t^{n+1}$$

having suitable properties. Perron was interested only in regular methods, and showed that the conditions

$$c_n \geqq 0 \quad (n = 0, 1, 2, \ldots), \tag{a}$$

$$f(1) = 1 \tag{b}$$

are sufficient for regularity. In order to prove other results, he introduced the further condition·

> The series $\Sigma c_n t^{n+1}$ has a radius of convergence greater than 1, and $z = f(t)$ maps $|t| = 1$ into a simple closed curved J which is cut by each half-ray from the origin in one point only. Consequently (Titchmarsh, p. 101), $z = f(t)$ maps $|t| < 1$ simply on D, the interior of J. (c)

Our present purpose is to continue the study of this summation method, but instead of requiring regularity, we shall require merely that the method be consistent with convergence. For this reason we drop condition (a) and replace it by the new condition:

> The curve J is not tangent to the real axis at $z = 1$. (d)

It will be shown that this, taken with (b) and (c), is a sufficient condition for the associated summation method to be consistent with convergence. For convenience we now make the ***permanent assumption*** that the function $f(t) = \Sigma c_n t^{n+1}$ satisfies conditions (b), (c), (d), and provide an explicit statement of the summation method:

Given a series Σa_n, write down formally the series $\Sigma a_n [f(t)]^{n+1}$, and arrange it in powers of t; let this give $\Sigma b_n t^{n+1}$. We call Σb_n the f-transform of Σa_n, and if $\Sigma b_n = S$, we say that Σa_n is f-summable to S.

To make the nature of this rearrangement clear, we put

$$[f(t)]^{n+1} = \Sigma_{m=n}^{\infty} c_m^{(n+1)} t^{m+1},$$

and write it out as follows:

$$\begin{array}{rcllll} a_0 f(t) & = & a_0 c_0 t + & a_0 c_1 t^2 & + \; a_0 c_2 t^3 & + \dots \\ a_1 [\, f(t)\,]^2 & = & & a_1 c_1^{(2)} t^2 & + \; a_1 c_2^{(2)} t^3 & + \dots \\ a_2 [\, f(t)\,]^3 & = & & & a_2 c_2^{(3)} t^3 & + \dots \end{array} \qquad \text{(A)}$$

$$\Sigma a_n [\, f(t)\,]^{n+1} = b_0 t + b_1 t^2 + b_2 t^3 +$$

We see that the method is essentially a matrix transformation.

THEOREM 1. *If Σa_n is f-summable, then $\Sigma a_n z^{n+1}$ has positive radius of convergence, and represents a function which can be continued analytically through D, the image of $|t| < 1$ by the function $z = f(t)$.*

PROOF. If Σb_n converges, then $\Sigma b_n t^{n+1}$ represents an analytic function for $|t| < 1$. Denote it by $\phi(t)$. Let the inverse of $z = f(t)$ be $t = g(z)$. Then the function $\psi(z) = \phi[g(z)]$ is analytic on D and can be represented by a power series $\Sigma a_n z^{n+1}$ in a circle with centre at the origin and lying in D. But $\psi[f(t)] = \phi(t)$, and so $\Sigma a_n[f(t)]^{n+1} = \Sigma b_n t^{n+1}$. It can be seen by a glance at (A) that $\alpha_0 = a_0$, $\alpha_1 = a_1$, and so on. The theorem now follows.

THEOREM 2. *If $\Sigma a_n z^{n+1}$ can be continued analytically through D', the closure of D, then Σa_n is f-summable.*

PROOF. We see at once that $\phi(t)$ is analytic for $|t| \leqq 1$, and therefore in particular that $\Sigma b_n t^{n+1}$ converges with $t = 1$. This proves the theorem.

We shall use the notation $\phi(t) = \Sigma b_n t^{n+1}$, $\psi(z) = \Sigma a_n z^{n+1}$ throughout this paper. We shall also use D, D' as defined in Theorems 1 and 2.

These two theorems, taken together, closely resemble Theorems 6 and 7 of Perron. We wish, however, to point out their bearing on $E(r)$ summability. In accordance with the opening paragraph of this paper, we put

$$z = \frac{rt}{1-(1-r)t} \qquad (|1-r| < 1).$$

This gives

$$t = \frac{z}{r+(1-r)z},$$

and the circle $|t| = 1$ maps into the circle

$$\mathfrak{S}: \qquad |z| = |r+(1-r)z|,$$

which gives at once the second part of Theorem 8.4 of Agnew, namely, that if Σa_n is summable $E(r)$, then $\Sigma a_n z^{n+1}$ represents a function analytic inside $\mathfrak{S}$. (Since $|1-r| < 1$, $E(r)$ and $\mathfrak{E}(r)$ are equivalent (Agnew, p. 327).) At the same time we see that the theorem is "best possible" in the sense that no more is true in general, and we get the complementary theorem:

THEOREM 2a. *If $\Sigma a_n z^{n+1}$ represents a function which can be continued analytically through the region $|z| \leqq |r+(1-r)z|$, then Σa_n is summable $E(r)$, provided $|1-r| < 1$.*

The circle $\mathfrak{S}$ can be written

$$|z| = |r-1|\left|z - \frac{r}{r-1}\right|.$$

It is easily verified (see Titchmarsh, p. 191) that for any values of p, q, and real positive $k \neq 1$, the circle

$$|z-p| = k|z-q|$$

has centre

$$z_1 = \frac{p - k^2 q}{1 - k^2}$$

and radius

$$\rho_1 = \frac{k|p - q|}{|1 - k^2|}.$$

Hence $\mathfrak{J}$ is a circle with centre

$$z_1 = \frac{|1 - r|^2}{1 - |1 - r|^2} \frac{r}{1 - r}$$

and radius

$$\rho_1 = \frac{|r|}{1 - |1 - r|^2}.$$

The shortest distance from the origin to this circle is given by

$$\rho_1 - |z_1| = \frac{|r|}{1 - |1 - r|^2} - \frac{|1 - r|\,|r|}{1 - |1 - r|^2}$$

$$= \frac{|r|}{1 + |1 - r|} = \frac{1}{\left|\frac{1}{r}\right| + \left|1 - \frac{1}{r}\right|}.$$

This gives the radius of convergence of $\Sigma a_n z^{n+1}$, in agreement with Agnew's "necessary condition for summability $E(r)$" (p. 315).

Furthermore, if $R > 1$, a necessary and sufficient condition for $f(t)$ to have the property that Σa_n is f-summable whenever $\Sigma a_n z^{n+1}$ has its radius of convergence greater than or equal to R, is that D' should lie in the circle $|z| < R$. (If $R \leqq 1$, the last requirement contradicts condition (b).)

For $E(r)$ summability, the condition is $\rho_1 + |z_1| < R$, or

$$\frac{|r|}{1 - |1 - r|} < R,$$

or

$$|r/R| + |1 - r| < 1.$$

Thus we have generalized a theorem of the author's (Macphail, 1946).

THEOREM 3. *If Σa_n is f-summable to S, it is "generalized Abel" summable; that is, $\psi(z) \to S$ as $z \to 1-$ on the real axis.*

PROOF. Let L be the segment of the real axis joining $z = 0$ to $z = 1$, and let C be the image of L in the t-plane. By condition (d), L is not tangent to J; hence C is not tangent to $|t| = 1$. Hence by Abel's theorem, $\phi(t) \to S$ as $t \to 1-$ on C. Since $\psi(z) = \phi(t)$, the theorem now follows.

COROLLARY. *Any two f-summation methods are consistent; in particular, any f-summation method is consistent with convergence.*

THEOREM 4. *Let* Σa_n *be f-summable. Then* $\Sigma a_n w^{n+1}$ *is f-summable for all values of* w *such that* $wD' \subset D$ *(that is,* z *in* D' *implies* wz *in* D*).*

PROOF. If we write $\Psi(z) = \Sigma a_n w^{n+1} z^{n+1}$, a sufficient condition for $\Sigma a_n w^{n+1}$ to be f-summable is that $\Psi(z)$ be analytic for z in D' (Theorem 2). But $\Psi(z) = \psi(wz)$, and since ψ is known to be analytic in D, the result follows.

For the application to $E(r)$, we note as before that D' is a circle with centre

$$z_1 = \frac{|1 - r|^2}{1 - |1 - r|^2} \frac{r}{1 - r}$$

and radius

$$\rho_1 = \frac{|r|}{1 - |1 - r|^2}.$$

Then wD' is a circle with centre $z' = z_1 w$, and radius $\rho' = \rho_1 |w|$. For wD' to lie in D, we require

$$|z' - z_1| + \rho' < \rho_1,$$

whence,

$$\frac{|1 - r|\,|r|\,|1 - w|}{1 - |1 - r|^2} + \frac{|r|\,|w|}{1 - |1 - r|^2} < \frac{|r|}{1 - |1 - r|^2},$$

or,

$$|r - 1|\,|w - 1| + |w| < 1,$$

which gives the first part of Theorem 8.4 of Agnew. Here again we see that the theorem is the best possible.

It should be mentioned at this point that one of the main results of the theory, already proved by Knopp, Perron, and Agnew, holds without change under our assumptions (b), (c), (d). This result enables us to say exactly in what region a given series is f-summable, with doubt only on the boundary of the region. For completeness we shall state it here, and in order to do so we require the following definitions.

Let $\psi(w) = \Sigma a_n w^{n+1}$ be a given series, and let $\psi^*(w)$ be the analytic continuation of $\psi(w)$ throughout the associated Mittag-Leffler star $\mathfrak{S}$. Let ω be any vertex of $\mathfrak{S}$, and J_ω the image of $|t| = 1$ by the function

$$w = \frac{\omega}{f(t)}.$$

Let I_ω, E_ω, be the interior and exterior of J_ω, and let I be the intersection of the I_ω, E the union of the E_ω, for all vertices ω.

THEOREM 5. *The series $\Sigma a_n w^{n-1}$ is f-summable to $\psi(w)$ if w is in I but not if w is in E.*

The proof given by Perron (Theorem 9) holds under our present assumptions (b), (c), (d), in fact under (b), (c) alone, and is not repeated here.

In applying this to $E(r)$ summability, we write simply

$$w = \frac{\omega[1 - (1 - r)t]}{rt},$$

whence

$$t = \frac{\omega}{rw + \omega - \omega r},$$

and the image of $|t| = 1$ is

$$|rw + \omega - \omega r| = |\omega|,$$

or,

$$\left| w - \left(1 - \frac{1}{r}\right)\omega \right| = \left| \frac{\omega}{r} \right|.$$

This is the boundary of the region $B(r, \omega)$ of Agnew (p. 328), and the theorem just stated is the extension of Theorem 8.2 of Agnew.

In conclusion we state two theorems where, instead of assuming that Σa_n is f-summable, i.e., that Σb_n converges, we assume that $\Sigma b_n t^{n+1}$ converges for $|t| < h$. This is a stronger or weaker hypothesis according as h is greater or less than 1. Let D_h be the image of $|t| < h$ by the function $z = f(t) = \Sigma c_n t^{n+1}$, and assume that the analogue of condition (c) holds.

THEOREM 6. *Let $\Sigma b_n t^{n+1}$ converge for $|t| < h$. Then $\Sigma a_n z^{n+1}$ is f-summable for all values of z such that $zD' \subset D_h$, and represents an analytic function which can be continued through D_h.*

The proof is omitted, being practically identical with that of Theorems 1 and 4.

For $\mathfrak{E}(r)$ summability, we find that D_h is the circle with centre

$$z_h = \frac{h^2|1 - r|^2}{1 - h^2|1 - r|^2} \frac{r}{1 - r},$$

and radius

$$\rho_h = \frac{h|r|}{1 - h^2|1 - r|^2},$$

the restriction $h|1 - r| < 1$ being necessary. By a little calculation we can also establish the following result:

Theorem 7. *Let the* $\mathfrak{E}(r)$ *transform of* Σa_n *be* ΣA_n. *If* $\Sigma A_n t^{n+1}$ *converges for* $|t| < h$, *then* $\Sigma a_n z^{n+1}$ *converges for*

$$|z| < \frac{h|r|}{1 + h|1 - r|}.$$

and $\Sigma a_n z^{n+1}$ *is summable* $\mathfrak{E}(r)$ *for all values of* z *satisfying*

$$|r - 1|\,|z - A| + |z| < B,$$

where

$$B = h\frac{1 - |1 - r|^2}{1 - h^2|1 - r|^2}, \quad A = hB.$$

Here again we omit the proof, which follows closely the pattern of the remarks following Theorems 2 and 4.

Queen's University

REFERENCES

Agnew, R. P. (1944). Euler transformations. Amer. J. Math. **66**: 313-38.

Knopp, K. (1922, 1923). Uber das Eulersche Summierungsverfahren (I, II). Mathematische Zeitschrift **15**. 226-53; **18**: 125-56.

Macphail, M. S (1946). Euler-Knopp summability of classes of convergent series. Amer. J. Math. **68**: 419-50.

Perron, O. (1923) Uber eine Verallgemeinerung der Eulerschen Reihentransformation. Math. Zeit. **18**: 157-62.

Titchmarsh, E. C. (1932). The theory of functions. Oxford.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLII : SERIES III . MAY, 1948
SECTION THREE

The Stability of Gas Turbine Combustors

By DONALD L. MORDELL

Presented by A. NORMAN SHAW, F.R.S.C.

INTRODUCTION

The use of the gas turbine engine for aircraft propulsion necessitates a combustion system capable of efficient operation over a range of temperature rises between 10 and 1000° C. in spite of wide variations in the pressure, temperature, and quantity of the air to be heated. Such performance, referred to as stability, is of the utmost importance, because in its absence it may be impossible to operate the engine above certain critical altitudes, or in certain ranges of engine speed—limitations which are obviously undesirable in an aircraft engine.

Systems to meet these requirements have been developed by empirical means, and it is of interest that the solutions evolved independently in the United Kingdom, the United States, and Germany are basically similar. The features common to all may be described in a few words. The air stream to be heated is divided at entry. About 25 per cent is used for primary combustion, the remaining 75 per cent being mixed afterwards with the hot combustion products. The primary air is mixed with the fuel, which is injected either as a vapour or as an atomized spray, through the agency of a complex flow pattern.

The functions of this pattern are to provide turbulence to promote mixing of the air and fuel, to spread around burning fuel particles, which then act as torches to help enkindle other particles. Also in all designs there is what may be termed a major flow reversal which recirculates heat from the burnt fuel to the incoming fuel air mixture by a reverse flow of hot gases. This reverse flow of heat is also aided by direct radiation.

In known combustors, it has been found that there are definite limits to the temperature rise at which the system will operate, and within these limits, definite values of the thermal or combustion efficiency. This is defined as the ratio of the heat actually liberated to the heat input in the fuel. The stability limits are functions of the geometry of the combustor and the pressure, temperature, and velocity of the air (Childs, McCafferty, and Surine).

"IDEAL" COMBUSTOR

In attempting to explain this behaviour it is evident that any exact treatment is out of the question, involving as it does the aerodynamics of turbulence, evaporation of drops, the chemistry of combustion, and radiation.

The method used by the author is to postulate an ideal combustor in which there is a separation of the complex interrelated processes into phases which are more amenable to experimental study. Then from results of tests on these phases, the performance of the ideal combustor may be synthesized and compared with that of the actual combustors.

Consideration of the real combustors suggests that the predominant mechanism is a recirculation of heat from the burnt products to the incoming air. The fuel may be considered as being injected into air thus heated, and if the temperature be sufficient, spontaneous ignition will occur.

For the ideal combustor, we suppose first a division of air into primary and dilution streams. The primary air enters the actual combustion region and is mixed with a quantity of hot gases which have been recirculated from the end of the primary combustion system, and are presumed to be at the temperature existing there. Also at the end of the primary combustion region, the dilution air is injected and the resultant drop in temperature is assumed to arrest any further chemical reaction.

We have thus replaced the complex aerodynamic and heat transfer system by a scheme which can be specified by three factors: a scale factor, the proportion of primary air, and the proportion of air participating in the reversal. For the combustion side we have reduced our problem to the question of what happens when fuel is injected into a hot gas stream.

P. Lloyd (1946 *a*, *b*) has shown that, in such a case, a stable combustion region is created some distance downstream of the point of fuel injection. The distance between this point and the flame front may be interpreted in terms of a time, called the ignition delay, and related to the pressure and temperature of the air by a relation of the type:

$$\log t = \log \frac{14.7}{P} + A + \frac{B}{T}$$

where P is the pressure in p.s.i., for $P < 14.7$, T is the temperature, A and B are constants. So far, Lloyd's work has shown that A and B are practically independent of the mixture strength and the chemical nature of the fuel. Lloyd's experiment provides a simple way of studying fuel air mixing and combustion that gives results in a method convenient for analysis. As yet no results have been reported giving the progress of combustion beyond the flame front, i.e., the time required for completion of combustion after the initial ignition—defined as the appearance of a definite flame front.

In the ideal chamber, the time during which any given particle of air remains in the combustion region, i.e. between heating by the recirculating hot gases and quenching by the admixture of the dilution air, is obviously a relatively simple function of the initial volume flow of air, the three factors specifying the aerodynamic system, and the temperature following primary combustion. Since Lloyd's experiment provides a means of studying combustion on a time basis, we can conveniently use the results to predict the behaviour of our ideal combustor and in particular to calculate how the range of temperature rises and combustion efficiency will depend upon the pressure, temperature, and rate of the air flow, and upon the chamber geometry.

APPLICATION TO LOWER LIMITING TEMPERATURE RISE

In any combustion system, as the air fuel ratio is gradually increased a point comes when the flame extinguishes. The point may be defined as the maximum air fuel ratio, but, in the author's view, it is more convenient and of greater significance to define it as the lower limiting temperature rise.

In the ideal chamber it is evident that combustion is impossible when any particle of air is within the primary region for a shorter length of time than the required ignition delay. At the limiting point of extinction the times may be equated, yielding a relation of the type

$$\log \frac{K_1}{V\left(\frac{T_2}{T_1} - K_2\right)} = A + \log \frac{14.7}{P_1} + B \frac{K_3}{T_1\left(\frac{T_2}{T_1} - K_2\right)}$$

wherein K_1, K_2, and K_3 are functions of the three factors specifying the aerodynamics of the system, A and B are constants representing the physics of droplet evaporation and the chemistry of combustion, V, P_1, and T_1 are the rate of volume flow, the pressure and the temperature of the entry air stream, and T_2 is the limiting, minimum outlet temperature.

If the effects of V, P_1, and T_1 upon the limiting rise $T_2 - T_1$ are calculated, it is found that the ideal chamber behaves qualitatively exactly as a real combustor. A particular point of interest is that it has been found that change of fuel in a real combustor, e.g. from gasoline to kerosene, affects the performance considerably in many respects; however, it does not affect the lower limiting temperature rise. In predicting the effect of fuel change on the ideal chamber, Lloyd's work shows no significant change in A or B with the change of fuel, and hence no effect on the lower limiting rise is to be anticipated.

DISCUSSION

The analysis has been extended to other features and aspects of performance and so far has always yielded results in qualitative agreement with test. Disagreement in numerical values is believed to lie principally in the idealisation of the mode of heat recirculation. Work presently in hand in the Mechanical Engineering laboratory at McGill University is concerned with further exploration of this point, with a study of the progress of combustion after initial ignition and the effect of turbulence upon the delay and subsequent combustion. A more detailed paper giving the full treatment using these results will be published in due course. The reason for the publication of this preliminary outline is that it is thought to be the first attempt at an explanation of the very complex and sometimes apparently paradoxical results which have been found in practice.

REFERENCES

CHILDS, J. H., MCCAFFERTY, R. J., and SURINE, O. W. (1947). Effect of combustor inlet conditions on performance of an annular chamber. Washington, National Advisory Council for Aeronautics, Technical note no. 1357.

LLOYD, P. (1946). Combustion in the gas turbine. London, Institution of Mechanical Engineers, War Emergency issue no. 12.

——— (1946). Spontaneous ignition of liquid fuels in a hot gas stream. Paris, Applied Mechanics Congress.

TRANSACTIONS OF

THE ROYAL SOCIETY OF CANADA

SECTION III

CHEMICAL, MATHEMATICAL, AND PHYSICAL SCIENCES

THIRD SERIES—VOLUME XLIII—SECTION III

JUNE, 1949

56671

Ottawa
The Royal Society of Canada
1949

TABLE OF CONTENTS

A Quarter-Century of Research in Physics. Presidential Address, by J. S. Foster, F.R.S.C. 1

The Mellin Transform of Functions Defined by Infinite Series. By R. M. Goodspeed 15

The Limit Points of Riemann Sums. By R. L. Jeffery, F.R.S.C. and C. N. Rowse 21

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLIII : SERIES III : JUNE, 1949
SECTION THREE

PRESIDENTIAL ADDRESS

A Quarter-Century of Research in Physics

By J. S. FOSTER, F.R.S.C.

IT is a great honour and a real pleasure to address you here in my native province and especially at Dalhousie—the university home of an appreciable portion of Canada's first class scientists.

Thanks to the understanding and wisdom of our government in this free country, certain novel and effective developments were made by scientists during the war, just past, and mainly as a result of this demonstration of strength, we find a great increase in the appreciation of science by the public at large and by university, provincial, and federal authorities. This appreciation is gradually finding the expression most valued by disciplined scientists, namely an opportunity to increase the scope and quality of scientific work. With the excessive post-war burden of teaching now becoming lighter, the universities find a rising scientific spirit which is nourished by thoughts put into action.

The inhabitants of this province have always been intensely interested in education. The Nova Scotia Institute of Science existed many years before the formation of the Royal Society of Canada. With this background, it is therefore not surprising to note that the half-million people of this province have set up a Nova Scotia Research Foundation which has already proved a great stimulus to scientific effort. Nor is it surprising that the sound policies in scientific work have been further rewarded in this old city by an even greater federal development. On behalf of this Section of The Royal Society of Canada, I extend to local scientists our congratulations on the new order of magnitude of scientific opportunities thus afforded and express our best wishes for their continued success.

I now take this opportunity to review briefly the researches in which I have been chiefly interested since coming to McGill University twenty-five years ago.

Introduction to the Stark Effect

At the beginning of this period, as you all know, physicists were chiefly interested in the structure and behaviour of the outer atom.

Studies of atoms in external fields, both electric and magnetic, contributed much to the general development. The normal Zeeman effect had immediately received an (imperfect) classical explanation but gave so little hope for an appreciable effect from external electric fields—arising only from departure from Hooke's law—that Zeeman tried to combine the two fields and detect the electric phenomena as a small perturbation on the normal Zeeman effect.

The point of view was completely changed by the Bohr frequency condition, and an appreciable electric effect was in fact clearly predicted by Bohr early in 1913. Large displacements observed in hydrogen by Stark in the following year and a half received a quantum theoretical explanation by Epstein who used parabolic coordinates in the Hamilton-Jacobi equation. This was an extension to three dimensions of a method developed by Sommerfeld to explain fine structure in hydrogen, but nevertheless it failed to represent such structure, since no variables have been found which are separable in this equation when both external electric field and relativistic corrections are represented.

To speak now of the experimental side, a rather difficult problem arises. For best results, one needs to keep the distribution of space charge near the cathode of a discharge tube of small diameter, constant over a period of hours while the analysis is being made by a spectrograph of high dispersion. Under high voltage, a beam of positive ions is driven through the so-called cathode dark space and here—along the axis—atoms are excited to give relatively intense light. Electric fields ranging from zero at the upper limit of the dark space to a maximum near the cathode are created and maintained by space charges.

After an exhaustive survey of experimental methods, the following technique solved the problem for many gas discharges. A lavite cylinder about a half-inch in diameter was bored to receive a heavy cathode exposed to the discharge through a small hole off centre, while light was removed through a very narrow slit and out through a side tube. Since the lavite may be machined with precision before heat treatment and since it seals into pyrex glass to prevent side discharges, the control over the discharge was immensely improved, and difficulties due to sputtering removed. The aluminum cathode was pitted by the positive ions, but rotation of the cathode exposed a fresh surface. This basic development of experimental method and preliminary results were realized during the tenure of a National Research Fellowship under Rockefeller funds.

VALUE OF FELLOWSHIPS

I pause to remark that this is only an illustration of the value of a Fellowship to students who have already received their Ph.D. degrees. Few students are thoroughly established in a new field by the time they graduate, and those who look forward to intensive research need a few years to get well grounded while the demands of life are relatively simple. While it is obvious that such a plan might change the whole life of the individuals concerned, one has only to recall such Fellowships held by Professors Compton, Lawrence, Rabi, Urey, etc., to realize that the scientific life on this continent has been profoundly changed by what must now be regarded as a very modest investment.

STARK EFFECT A STIMULUS TO IMPORTANT THEORETICAL DEVELOPMENT

Returning now to the interesting interplay between experiment and theory in the development of this phenomena so characteristic of quantum theory, I may say that the early results in helium by improved experimental techniques were used even before publication by Kramers and Heisenberg, in an attempt to apply perturbation theory to the Stark effect in helium. The results of this investigation, which was never published, were always qualitatively satisfactory, and quantitatively incorrect. As a result of this exercise, Heisenberg became convinced that one must abandon the whole mathematical method then used. By a bold stroke, he concluded that instead of the *harmonic* analysis of electron orbits—long known to be so foreign to atomic behaviour—one must accept only the frequencies and amplitudes characteristic of the light from the system under investigation. By such methods he considered the simple harmonic oscillator and arrived at an expression for the energy already known to be necessary in the explanation of molecular spectra, namely $E = (n + \frac{1}{2})h\nu$, and quantum mechanics was established.

DEGENERACY IN ELECTRIC FIELD

From elementary considerations, it is clear that the angular momentum of an electron round the nucleus could not remain constant in the presence of an external electric field, and under this condition, the ordinary selection rule for the quantum number l is removed and many new lines appear in the spectrum—corresponding to Δl equal to any possible value. These are the so-called "forbidden" lines, but since their appearance has been invited by the external electric field, it seems more logical to call them briefly, "electric combination lines." Now while the time mean position of an electron in its orbit will be drawn away from its normal position at the nucleus and thus change the energy of the system with respect to the field, it obviously makes no difference whether the electron is rotating right or left hand with respect to the field direction. This means that the sub-levels designated by the quantum numbers $\pm m$ will lie together and for values of m different from zero the system is doubly degenerate.

OPERATIONAL TECHNIQUE

In experimental work, it is not uncommon to find that the discharge tube suddenly becomes harder and the voltage increases abruptly so that by this means it is possible to manufacture spurious components. This is mentioned only because it actually happened in work by Takamine and such incorrect analyses in helium served as an unnatural barrier to any theoretical attempt to explain this phenomena.

While spurious components should not appear if one keeps a close eye on the voltmeter, one may be doubly sure of the analysis by watching the simultaneous analysis of hydrogen lines appearing on the same plate and noting that their pattern agrees with the Epstein theory.

NORMAL STARK PATTERNS

With this check, and generally improved means of analysis, the helium spectrum was examined for Stark patterns in the various series. There was immediate evidence that these were constant within a given spectral series, but varied from one series to another. Thus all principal and sharp series lines showed one component in each polarization while $2P$-nP combination lines showed two of each polarization. A few diffuse lines yielded two parallel and three perpendicular components. As a guide in a further examination of close line groups it was noticed that one could readily extend qualitatively to other hydrogen lines the perturbation theory by Kramers of connections between the fine structure and Stark components of H_α. When this was done, it was easily seen that the pattern should be constant within each fine structure series. These are indeed the patterns which were experimentally found in parhelium and which first established some law and order in the Stark effect.

The first pattern observed at McGill was for the $2S$-nS electric combination lines which according to the ordinary selection rule for the quantum number m could give only one component in which the electric vector is parallel to the field (π-component) and no component perpendicular to the field (σ-component). Thus a plane polarized series of spectral lines appeared in relatively high fields.

The positive identification of such lines in the ultra-violet is easily made in the following way. The initial levels in such transitions are alone appreciably affected by the field, and the relation between displacement of level and the applied electric field is commonly rather complex and characteristic of the individual level. Once these complex relations are established (e.g. in $2P$-nS) one can identify a new combination line (e.g. $2S$-nS) by its characteristic behaviour in the electric field. Thus the peculiar names assigned by Stark to some electric combination lines were eliminated on experimental as well as general theoretical grounds.

GENERAL FORMULAE FOR NORMAL STARK PATTERNS

Keeping in mind the degeneracy in Stark effect and the selection rule for m, one can readily write down the pattern as found for any series. If the l values of initial and final steps are equal, $l' = l''$, then the pattern will be represented by the following numbers of components:

$$\frac{\pi}{\sigma} = \frac{l+1}{2l}.$$

If $l' \neq l''$ then $\frac{\pi}{\sigma} = \frac{l+1}{2l+1}$ where the l value is the smaller of the two.

THEORY OF STARK EFFECT IN HELIUM

The quantum theoretical treatment of the Stark effect in helium is an ideal example of the perturbation theory in quantum mechanics and again emphasizes how essential the quantum theory is for this subject, since it is found that the displacements and intensities depend not on the behaviour of *single* states but on the *collective behaviour of certain groups* of sub-levels with the same principal quantum number, n. In this theory, a hydrogen atom is considered to be perturbed by (1) a central field which is asked to change the hydrogen terms into helium terms, and (2) an external electric field.

In the matrix H_1, the terms arising from the central field are simply constant terms on the diagonal and represent the differences between hydrogen and helium terms of same l value which are extracted from the tables or Lo Surdo photographs. The terms representing energy due to the external field are different from zero only when

$l' = l'' \neq 1$, and $m' = m''$. This means that levels with neighbouring l values and with the same m value repel one another. As a matter of fact, this repulsion is greater the less the separation between the levels and the more probable the transition between such levels in the normal atom. Due to this circumstance, one can divide the problem of finding the displacement of sub-levels into parts, considering first the level of highest m value which is never displaced, then the two levels of next highest m value which must be displaced equally in opposite directions and so on to the groups of sub-levels with $m = 0$ which as a rule will show some of the greatest displacements. It may be seen that this makes a great contrast with the Zeeman effect where each level with $m = 0$ is not displaced and where the levels with the maximum = m values have the greatest displacements in opposite directions. The intensities are closely tied up with displacements, so that one may say, for example, that an electric combination line will not appear until the external electric field has displaced the initial level slightly. Since such displacements with rising field will be delayed for lines greatly separated from the diffuse or principal series lines appearing in the normal spectrum, it may be seen why the S-S series appears only in high fields while many other new series appear very close to the lines of the normal spectrum at almost zero field. The complete theoretical calculations were carried out in Copenhagen under an International Education Board Fellowship.

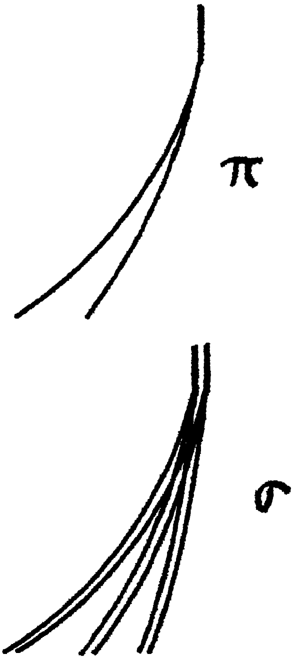
Before leaving this subject we may emphasize again that in the normal effects, i.e., in cases where the electron spin may be neglected, there is but one normal Zeeman effect for all lines of all series whereas the normal Stark effect shows a great variety of patterns as previously indicated.

INTENSITIES IN HYDROGEN

At McGill we initially obtained hydrogen lines from an impurity of hydrogen in helium discharge tubes. Under such conditions we never could get the relative intensities of Stark components in hydrogen to correspond to the results obtained by Stark in the canal-ray tube. Nevertheless when the new calculations were made by Schroedinger, using wave mechanics, it was immediately evident and quantitatively checked that the new theoretical intensities within experimental error were exactly as observed by Foster and Chalk. On the other hand, even in the same type of discharge tube which we had developed, if one used pure hydrogen one got an asymmetry observed by McRae for H_α. Moreover, the displacements were never exactly according to Epstein, but were modified just sufficiently to account for the original fine structure separations from which the components emerged. A further variation in intensities was observed in the Balmer series by Foster and Snell, who used a mixture of pure hydrogen and deuterium in equal parts; it consisted of right and left hand asymmetries for hydrogen and deuterium patterns of one and the same Balmer line. Since theoretical intensities have already been found from sources containing only a *trace* of hydrogen, it seems beyond doubt that the variations from the theoretical intensity pattern—such as those just mentioned—are due to self-absorption within a source containing relatively large numbers of polarized atoms of hydrogen and deuterium.

Effects due to Parallel Electric and Magnetic Fields

If one threads a discharge tube of small diameter along the axis of an electromagnet and through small holes in the pole pieces, then parallel electric and magnetic fields may be applied to atoms near the cathode which is suitably placed between the poles. In parhelium, the effects of the two fields are strictly additive with the result that π Stark components are unaffected while σ components are each split into two components with normal Zeeman separation. This is illustrated in the above diagram for a typical diffuse series line.



It is interesting to note that as the electric field goes to zero, three components will collapse into each of the σ components of the normal Zeeman effect for this line and two into the π component. Of the latter, one is doubly degenerate. In this way it is possible to see before your eyes a complexity in normal Zeeman effect not explained in the classical theory which is still used on occasions.

EFFECTS DUE TO PERPENDICULAR ELECTRIC AND MAGNETIC FIELDS

The problem of crossed electric and magnetic fields applied to atoms appeared at first to be one of exceptional difficulty in both experiment and theory. A magnetic field applied at right angles to a discharge tube would obviously put out the light; in early theory some of the components under such conditions were thought to be diffuse. The experimental problem was solved at McGill in stages and with different arrangements. I shall here recall only the final form used by Foster and Pounder. A small canal-ray tube is set up along the axis of a magnet in which the pole pieces themselves serve as cathodes and are bored with a 1/16 in. diameter hole so that canal rays shoot in opposite directions through the space between the poles. Here electrostatic plates are set up to provide the electric field at right angles to the magnetic and observations are made along the third axis at right angles to each field. This experimental arrangement gave a magnetic field of 25,800 gauss and electric fields up to 120,000 volts per cm. The photographs showed much more complex structures than normally found in Stark effect—with magnetic separations varying considerably each way from the normal separation found with parallel fields. They also showed pronounced changes in intensity with increasing electric field and in at least one case a striking variation in displacement with small change in the electric field. One of the more novel features was the appearance of three components arising from transitions in which $\Delta m = \pm 2$. It will be realized that the quantum number m can no longer have the physical significance which it has in either field separately or in parallel fields—for it certainly cannot represent the component of angular momentum along each of two fields at right angles. It becomes simply an index in the matrix theory.

THEORY OF CROSSED FIELDS

The theoretical treatment for such a case now becomes more difficult since *all* sub-levels with principal quantum number n must be considered at one time. The theory has been worked out in Professor Heisenberg's Institute at Leipzig before the war, but has not been published. Insofar as one can judge from the observations over range of fields overlapping in experiments and theory, there appears to be very good agreement.

STARK EFFECT IN COMPLEX SPECTRA

The McGill group examined the Stark effect in the remaining rare gases with the exception of radon and in mercury and a few other elements. The point about complex spectra is that while the external electric field exerts a force on the *charge* of the electron it has no direct effect on the *spin*. However, the magnetic field induced by the spin will be coupled to the magnetic field provided by the electron in its orbit and the resulting complications are understood

POLE EFFECT IS NOT A STARK EFFECT

Special reference may be made to iron where for many years scientists assumed that the small shift in the frequency of certain lines near the negative pole of an iron arc was in reality a Stark effect. Merely as a check on this, and to learn which lines were unreliable as standards, the Stark effect in iron was examined by Foster and Panter. We were surprised to learn that there is no apparent connection between the two effects, and indeed, some lines were shifted in the Stark effect in a direction opposite to that of their known shift in pole effect.

MOLECULAR SPECTRA

The law and order established in the Stark effect served as a stimulus to an important development in the theory of molecular spectra. Hund solved the following problem. A proton approaches a normal hydrogen atom and the external electric field from the proton produces a Stark effect in the hydrogen atom which increases as the proton approaches, and is continually changed into the molecular spectrum of the ionized hydrogen molecule as the proton takes its appropriate position. Indeed, as you know, the molecular terms receive their names from the components of angular momentum about the axis of the internal electric field of the molecule.

STARK EFFECT IN MOLECULAR SPECTRA

A theory of the Stark effect in the hydrogen molecule was worked out in general terms by Lorne MacDonald, 1851 Exhibition Scholar from McGill. MacDonald found that the effect depended upon a second order perturbation, but numerical comparison with experiment was prevented by a still imperfect knowledge of the normal molecular spectrum.

Experiments at McGill were carried out in the region H_α to H_β for the hydrogen molecule by Snell: for the deuterium molecule by Neamtan and for the HD molecule by Jones. A thirty-foot concave grating was employed in a stigmatic mounting. Over six hundred lines were examined. The patterns observed are rarely complete, but may be described by saying that the total angular momentum, K, is resolved to give $K + 1$ electric levels and that the polarization follows the usual rule.

The researches on the Stark effect which have just been reviewed are standard works in this field. They served incidentally as a very good training for about twenty-five graduate students, most of whom have never ceased to do research. Members of the group received an 1851 Exhibition Scholarship each year for six consecutive years and worked at Cambridge, London, Leipzig, Utrecht, and Berkeley. Some attention was paid to presentation of the work in publications. All papers, including theses, were accepted by the Royal Society of London in the original form.

QUANTITATIVE SPECTROGRAPHIC ANALYSIS

Overlapping somewhat with the investigation of Stark effects, a small group of us undertook to develop methods for the quantitative spectrographic analyses of trace elements in biological materials. This move was initiated by Dr. Weaver of the Rockefeller Foundation, who felt that spectrographic methods should be of value in medical research.

At the beginning, the main points in such work are, (1) to avoid contamination, and (2) to find a method which gives reproducable results. Such a method was developed with Drs. Langstroth and McRae. The essential point is to give each atom an equal chance of being excited and to this end the material was got into solution and put on a flat electrode with a buffer which helped to retain it. As a further precaution against loss of material, by evaporation, the sample was dried in a partial vacuum. By a mechanical device, the electrode was moved under a point and sparked evenly over the whole electrode. I must say that some alternative methods shock me considerably, but reproducibility and sensitivity are sufficient criteria for any method.

The quantitative feature was carried out by an internal standard and step-slit calibration which Langstroth had learned at Utrecht.

The technique needs to be varied in detail from element to element and one can rarely expect better than 5 per cent accuracy.

An alternative method for measuring intensities is the use of a neutral wedge in front of an evenly illuminated slit. In order to determine the difference in the lengths of lines so photographed, I used an optical system somewhat in the style of a range finder whereby the image of one line was moved relative to the other until they tapered off together. This device was found to give quite satisfactory results. It was used in a study with Horton of the boron deficiency in certain plants which leads to great loss. It was learned how late in the life of the plant boron could be added to save the crop, and the added boron could be traced through the plant in much the fashion that radioactive tracers are used.

In this connection, I should like to call attention to some work by Langstroth and McRae, in collaboration with Professor Babkin. It was known that the content of sodium and potassium in saliva could be changed by variation in the method of gland excitation. Not only were the quantitative analyses made under different types of excitation but Langstroth developed a theory by which one could predict the content of a sample taken at any time.

ULTRACENTRIFUGE

The problem of constructing and operating an ultracentrifuge proved particularly difficult until the methods of magnetic suspension and electrical drive devised by Beams were adopted. The absorption in a Svedberg cell was for the first time analysed by a quartz spectrograph attached to the system. This powerful tool was built up largely through the efforts of Drs. Rotenberg, Gunton, and Carruthers.

RESEARCH DURING WAR

In the first year of the war, the McGill group did succeed in making the first automatic target-following device in Canada and a range finder which would show on a dial the range of a moving target. Professor D. K. Froman was a member of the group—also Professor W. H. Watson, who later, in collaboration with Professor Terroux, Dr. Guptill and others, developed an exceedingly good resonant slot array. Since it seemed difficult, however, to find any arrangement whereby physical research done in universities stood much chance of being used in the war, I spent the following three years at the Radiation Laboratory, M.I.T. Here it was found possible to devise suitable wave guide feeds for antennae and a device to determine phase, and plot wave fronts, so as to know the origin of waves from such feeds. Later a type of radar scanner was invented which I shall describe briefly, mainly because it represents a translation of extremely simple ideas into an effective device.

If one stands in a parallel beam of radiation with a reflecting mirror then by rotation of the mirror the reflected beam may be swept over a limited angular range. The solution to be described uses this principle to produce a saw tooth scan of a parallel beam through a pre-determined angle at high speeds by steady rotation. In practice two coaxial conical conducting surfaces are set up with small separation. The inner one is rotated while the outer one is stationary. By a suitable device, a parallel beam is emitted from within the inner surface and emerges across an element of the cone to travel round the space between the two surfaces. A barrier attached to the outer surface picks off the radiation and takes it out to supply a line source at the focus of a parabolic cylinder. If one develops the conical scanning space on a plane one sees that in the best arrangement the wave fronts bisect the angle of the sector and that the barrier (mirror) which picks up the radiation rotates uniformly about them. On looking into this one finds (1) that owing to the simplest optics with a parallel beam, there is no coma or other optical troubles; (2) the device is insensitive to frequency; (3) the length of the wave front is the same throughout the scan with the result that side lobes do not change; (4) owing to the simple rotation, the scan can be very fast and the presentation gives a motion picture of proceedings on an ordinary cathode-ray tube.

By actual test this device had a higher efficiency than any other comparable scanner made during the war. The principle can be used in a great variety of applications and the development is still going on at McGill.

Radiation Laboratory and Cyclotron

In the fall of 1937 the Governors of McGill University voted money for a Radiation Laboratory and cyclotron, and the following two years were spent on the plans with the help of Dr. McRae and Professor Terroux. The war broke out just as the plans were complete. The university authorities considered it unwise to proceed with the project. Dr. McRae, who was a highly trained man and so-to-speak the personification of intellectual integrity, died suddenly.

Towards the end of the war in which the value of nuclear research was amply demonstrated, the project was revived by the McGill Governors on a much larger scale. The plans for the building and for the cyclotron were completely changed and developed with great care to ensure a well-balanced unit. The Laboratory, designed with the help of Professor Terroux, is a complete unit built into a side hill so that the cyclotron room may be covered with at least ten feet of earth for adequate protection, while the three-storey laboratory itself is well lighted. The connection between the two is by an underground tunnel with two right-angle bends to prevent radiation from entering the laboratory. Here on three floors there is very attractive laboratory space for twenty-seven scientific workers as well as machine shop, seminar room, glass blowing room, dark room, and a penthouse for the radar development. Good ventilation is provided for certain rooms including a very good chemical laboratory, built up by Dr. Lloyd Thompson, the glass blowing room, and a few rooms where the stronger sources of radioactivity may be handled.

The three hundred ton synchro-cyclotron uses a minimum of iron to get a proton beam of 100 Mev energy. This is excited by a grounded grid oscillator coupled to a half wavelength line. Distinctive features include magnet coils of extruded aluminum strip containing water cooling; relatively small magnet gap and thin dee electrode; top and bottom of square vacuum chamber welded directly to pole pieces; ceramic axle to keep radio frequency power away from bearings of variable condenser at end of half wavelength line.

The large vestibule entrance to cyclotron room—necessary for installation of magnet—is now closed, front and back, by water-tight doors and turned into a water tank in which automobile type radiators are immersed to cool water from the cyclotron by circulation in closed systems. The tank water itself is sprayed in an overhead cooling tower. Since the tank is twelve feet thick, the water provides good protection from neutrons.

For initial operation of a large cyclotron, it is essential to use a d.c. bias to sweep out ions as they appear above and below the dee—otherwise the oscillator is overloaded by the ions and fails. The McGill cyclotron ran on the first day of its trials—June 4, 1949—after a record of three hours spent in outgassing metal parts.

This research unit is not large, but it is exceptionally well balanced. It represents more capital investment than any other university has put into such an undertaking. We are therefore most grateful first and foremost to the Governors of McGill University. We have also to express our sincere thanks to the government officials and agencies for expressions of goodwill as well as for vital research equipment and assistance.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLIII : SERIES III : JUNE, 1949
SECTION THREE

The Mellin Transform of Functions Defined by Infinite Series

By F. M. GOODSPEED
Presented by R. L. JEFFERY

1. If the analytic function $\varphi(u)$ satisfies certain conditions, the series

$$\varphi(0) - x\varphi(1) + x^2\varphi(2) - \dots$$

converges for small values of x and defines a function $\Phi(x)$ which is regular for all real positive values of x. The Mellin transform of $\Phi(x)$ is then given by the formula

$$\int_0^\infty x^{s-1}\Phi(x)dx = \frac{\pi}{\sin \pi s}\ \varphi(-s)$$

which was stated by Ramanujan and proved by Hardy.[1]

The object of this paper is to extend this formula to include the case where the function $\Phi(x)$ is defined by a power series in x where the exponents are not necessarily integers.

2. The result that will be proven is the following, in which the notation $u = v + iw$ will be used.

Let the infinite sequence $\{\lambda_n\}$ *be such that* $0 < \lambda_1 < \lambda_2 \dots$ *and* $An \leq \lambda_n \leq Bn$ *where* $A > 0$ *and let*

$$C(u) = \prod_{n=1}^{\infty}\left(1 - \frac{u^2}{\lambda_n^2}\right). \tag{1}$$

Let $\varphi(u)$ *be regular, and let*

$$\varphi(v + iw) = O\,(e^{Pv - Q|w|}) \tag{2}$$

for $v \geq -\delta$ *where* $\delta > -\lambda_1$, *and* $Q < \pi/B$.

Then, if x *is real and small enough, some sub-sequence of the partial sums of the series*

$$-\sum_{k=1}^{\infty}\frac{\varphi(\lambda_k)}{C'(\lambda_k)}\ x^{\lambda_k} \tag{3}$$

converges to an analytic function $\Phi(x)$ *which is regular for all real positive* x *and*

---

[1] See *Ramanujan* by G. H. Hardy, Cambridge University Press, 1940 chap. XI.

$$(4) \qquad \int_0^\infty x^{s-1}\Phi(x)dx = \frac{\varphi(-s)}{C(s)}$$

for $-\lambda_1 < R(s) < \min(\lambda_1, \delta)$.

If the further condition

$$(5) \qquad \lim_{n\to\infty} (\lambda_{n+1} - \lambda_n) > 0$$

is imposed on the sequence $\{\lambda_n\}$, *then the series* (2) *converges in the ordinary sense for small* x.

The following information about the function $C(u)$ must first be derived:

(a) $C(u)$ is an integral function such that

$$(6) \qquad C(v + iw) > e^{\left(\frac{\pi}{B} - \epsilon\right)|u}$$

for any positive ϵ if $|w|$ is large enough and $|w| > |v|$.

$$(b) \qquad C(r_n + iw) > e^{-Kr_n}$$

for some constant K and for an unbounded sequence of real positive values r_n.

For (a), $\Sigma\, 1/\lambda_n^2 \leq \Sigma\, 1/A^2n^2$ and so the infinite product defining $C(u)$ is absolutely and uniformly convergent for all finite u and $C(u)$ is an integral function. Also, if $|w| > |v|$, then

$$\begin{aligned}
&|C(v + iw)| \\
&= \prod_{n=1}^{\infty} \left| 1 + \frac{w^2 - v^2}{\lambda_n^2} - 2i\frac{vw}{\lambda_n^2} \right| \geq \prod_{n=1}^{\infty} \left| 1 + \frac{w^2 - v^2}{B^2n^2} - 2i\frac{vw}{B^2n^2} \right| \\
&= \left| \frac{\sin\{\pi(v + iw)/B\}}{\pi(v + iw)/B} \right| > e^{\left(\frac{\pi}{B} - \epsilon\right)u}
\end{aligned}$$

for any positive ϵ if $|w|$ is sufficiently large.

In order to prove (b) it is sufficient to show that for fixed positive numbers F and G with $F < G$ and any positive number $V > V_0$ for some positive V_0, there is a number v in the interval (FV, GV) such that

$$C(v) > e^{-Kv}$$

where K does not depend on V or v. For if V takes the values $V_0 + 1$, $V_0 + 2$, $V_0 + 3$ . . . the corresponding values of v form the desired unbounded sequence since $|C(v + iw)| \geq |C(v)|$ for any w.

Let $FV \leq v \leq GV$ and let

$$\begin{aligned}
C(v) &= \prod_{n=1}^{\infty}\left(1 - \frac{v^2}{\lambda_n^2}\right) = \prod_{n=1}^{n_1-1} \cdot \prod_{n_1}^{n_2} \cdot \prod_{n_2+1}^{\infty} \left(1 - \frac{v^2}{\lambda_n^2}\right) \\
&= C_1(v)C_2(v)C_3(v)
\end{aligned}$$

where n_1 is the largest integer such that $Bn_1 \leq \frac{1}{2} F\Gamma$ and n_2 is the smallest integer such that $An_2 \geq 2\,G\Gamma$. Then for the range of n in $C_1(v)$

$$\frac{v}{\lambda_n} \geq \frac{F\Gamma}{Bn} > \frac{F\Gamma}{Bn_1} \geq 2$$

and

$$|C_1(v)| = \prod_{n=1}^{n_1-1} \left| 1 - \frac{v^2}{\lambda_n^2} \right| > \prod_{n=1}^{n_1-1} 3 > 1.$$

For the range of n in $C_3(v)$

$$\lambda_n^2 - v^2 \geq A^2n^2 - G^2\Gamma^2 \geq A^2n^2 - \tfrac{1}{4} A^2n_2^2 > \tfrac{3}{4} A^2n^2$$

and so

$$\frac{1}{C_3(v)} = \prod_{n_2-1}^{\infty} \left(1 + \frac{v^2}{\lambda_n^2 - v^2} \right)$$

$$< \prod_{n_2+1}^{\infty} \left(1 + \frac{4v^2}{3A^2n^2} \right) < \frac{\sin K_1 iv}{K_1 iv}$$

$$= O\left(e^{K_1 v} \right)$$

for some K_1 independent of Γ and v. Therefore

$$C_3(v) > De^{-K_1 v}.$$

If in $C_2(v)$ the transformation

$$v = \left(\frac{G+F}{2} + \frac{G-F}{2}\,\theta \right) \Gamma = (L + M\theta)\Gamma$$

is made, then

$$C_2(v) = \prod_{n=n_1}^{n_2} \left(1 - \frac{(L+M\theta)^2\Gamma^2}{\lambda_n^2} \right) = P(\theta)$$

where, corresponding to $F\Gamma \leq v \leq G\Gamma$, the range of θ is $-1 \leq \theta \leq +1$. Thus, $C_2(v) = P(\theta)$ is a polynomial in θ of degree d where

$$d < 2\left(\frac{2G\Gamma}{A} - \frac{F\Gamma}{2B} \right) + 6 < K_2\Gamma < K_3 v$$

for some K_2 and K_3 independent of Γ and v if $\Gamma > \Gamma_0$ for some constant Γ_0. If the leading coefficient of this polynomial is a, then

$$|a| = \prod_{n=n_1}^{n_2} \frac{M^2\Gamma^2}{\lambda_n^2} \geq M^d \prod_{n_1}^{n_2} \frac{\Gamma^2}{B^2n^2}$$

$$> M^d \prod_{n_1}^{n_2} \frac{\Gamma^2}{B^2n_2^2} > M^d \prod_{n_1}^{n_2} \frac{\Gamma^2}{B^2\{1 + (2G\Gamma/A)\}^2}$$

$$= M^d \prod_{n_1}^{n_2} \frac{A^2}{4B^2G^2\{1 + (A/2\Gamma G)\}^2}$$

$$> M^d \prod_{n_1}^{n_2} \left(\frac{A}{4BG} \right)^2 = \left\{ \frac{A(G-F)}{8BG} \right\}^d = M_1^d$$

for $\Gamma > A/2G$, if $M = (G-F)/2 \leq 1$.

Thus $P(\theta)$ is a polynomial in θ of degree $d < K_3 v$ with leading coefficient a where $|a| > M_1^d$.

But the properties of the Tchebycheff polynomials[2] show that a polynomial in θ of degree k with leading coefficient unity must attain an absolute value greater than or equal to 2^{-k+1} for some value of θ in the range $-1 \leq \theta \leq +1$.

Therefore, for some value of v in the interval (FV, GV)

$$\left|\frac{C_2(v)}{a}\right| = \left|\frac{P(\theta)}{a}\right| \geq 2^{-d-1}$$

or

$$|C_2(v)| > M_1^d 2^{-d-1} > 2(M_1/2)^{K_3 v}$$

since $M_1\ 2 = A(G-F)/16\,BG = \{A/B\}\,\{(G-F)/16\,G\} < 1$.

Collecting these results

$$|C(v)| > 2De^{-K_1}\,(M_1\ 2)^{K_3 v} > e^{-Kv}$$

for some constant K and some value of v in the range $FV \leq v \leq GV$ if $V > V_0$.

Now consider the integral

$$\frac{1}{2\pi i}\int \frac{\varphi(-u)}{C(u)}\,x^{-u}du$$

taken in the positive direction about the rectangular contour with corners $\sigma - iW$, $\sigma + iW$, $-r + iW$, $-r - iW$. Here $0 < r \neq \lambda_n$ for any n, $W > r$, $-\lambda_1 < \sigma < \min(\lambda_1, \delta)$, and x is real and positive.

The contributions to the value of the integral from the parts of the contour parallel to the real axis are

$$O \qquad {}^{-\left(\frac{\pi}{B}-\epsilon-Q\right)W}$$

as W tends to infinity, since $\varphi(-u) = O(e^{QW})$ and $C(u) > e^{\left(\frac{\pi}{B}-\epsilon\right)W}$ on these parts of the contour. But as $Q < \pi/B$ and ϵ can be as small as we please this tends to zero as W tends to infinity.

Therefore, applying the Cauchy residue theorem

$$\frac{1}{2\pi i}\int_{\delta - i\infty}^{\sigma + i\infty} \frac{\varphi(-u)}{C(u)}\,x^{-u}du = \sum_{\lambda_n < r} \frac{\varphi(\lambda_n)}{C'(-\lambda_n)}\,x^{\lambda_n} + \frac{1}{2\pi i}\int_{-r - i\infty}^{-r + i\infty} \frac{\varphi(-u)}{C(u)}\,x^{-u}du.$$

Now let r take one of the values r_n for which (b) is satisfied, and let

---

[2] See R. Courant and D. Hilbert, *Methoden der Mathematischen Physik*, Berlin, 1931, vol. I, p. 75

$$\int_{-r-i\infty}^{-r-i\infty} \frac{\varphi(-u)}{C(u)} x^{-u}du = \int_{-r-i\infty}^{-r-ir} + \int_{-r-ir}^{-r-ir} + \int_{-r-ir}^{-r-i\infty}$$
$$= I_1 + I_2 + I_3.$$

Then $I_2 = O(re^{Pr + Qr + Kr + r\log r})$ as $r \to \infty$ through the sequence of values r_n satisfying (b). Therefore, if $\log x < -(P + Q + K)$ or $x < e^{-(P+Q+K)}$, I_2 will tend to zero as $r \to \infty$ through this sequence. Also, for some constant E

$$|I_3| < E\int_r^{\infty} e^{\left\{Pr + Qu - \left(\frac{\pi}{B} - \epsilon\right)u + r\log r\right\}} dw$$
$$= Ee^{(P + \log r)r} \int_r^{\infty} e^{-\left(\frac{\pi}{B} - \epsilon - Q\right)u} dw.$$

Thus, since $Q < \pi/B$ and ϵ can be chosen so that $\epsilon < (\pi/B) - Q$, the integral in the last expression is bounded and $I_3 \to 0$ as $r \to \infty$ if $x < e^{-P}$.

Therefore, if x is real, positive, and small enough, the sequence

$$S_k = \sum_{\lambda_n < r_k} \frac{\varphi(\lambda_n)}{C'(-\lambda_n)} x^{\lambda_n} = -\sum_{\lambda_n < r_k} \frac{\varphi(\lambda_n)}{C'(\lambda_n)} x^{\lambda_n}$$

converges to the integral

$$\frac{1}{2\pi i} \int_{\sigma - i\infty}^{\sigma + i\infty} \frac{\varphi(-u)}{C(u)} x^{-u}du \tag{7}$$

as $k \to \infty$.

But if $x = \rho e^{i\phi}$, the integrand of this integral is

$$O\left(e^{-\left(\frac{\pi}{B} - Q - \epsilon\right)u - u\varphi}\right)$$

and so, arguing as before, the integral is absolutely and uniformly convergent for $|\varphi| < (\pi/B) - Q$ and any ρ. Therefore (7) represents an analytic function $\Phi(x)$, regular for $|\mathrm{am}\,x| < (\pi/B) - Q$ and, in particular, for x real and positive.

Finally, applying the Mellin inversion formula[3]

$$\int_0^{\infty} x^{s-1}\Phi(x)dx = \frac{\varphi(-s)}{C(s)}$$

for $-\lambda_1 < R(s) < \min(\lambda_1, s)$.

---

[3] E. C. Titchmarsh, *Fourier Integrals*, Oxford University Press, 1937, Theorem 29.

If the additional condition

$$\underline{\lim}_{n \to \infty} \quad (\lambda_{n+1} - \lambda_n) > 0$$

is imposed on $\{\lambda_n\}$, a general theorem[4] on functions of the type of $C(u)$ shows that $|C(v + iw)| > e^{-Kv}$ for some K and $v = (\lambda_{n+1} + \lambda_n)/2$ for all n. It then follows that the series (3) converges in the ordinary sense for x small enough.

---

[4] See V Bernstein, *Séries de Dirichlet*, Paris, 1933, p 276, Theorem 2 This theorem does not allow u to be real, but it is a trivial extension to include this case.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLIII : SERIES III : JUNE, 1949
SECTION THREE

The Limit Points of Riemann Sums

By R. L. JEFFERY, F.R.S.C. and C. V. ROWSE

LET $f(x)$ be a bounded function on the linear interval $[a, b]$. Let (x_{i-1}, x_i) be the intervals of a finite subdivision of $[a, b]$, and ξ any point with $x_{i-1} \leqq \xi_i \leqq x_i$. It has been shown by R. S. Sastry (1940) that as $n \to \infty$ the set of limit points of the sum

$$(1) \qquad \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1})$$

in every point of the closed interval $[m(b-a), M(b-a)]$ where m and M are respectively $\inf f(x)$, $\sup f(x)$, $a \leqq x \leqq b$. If in (1) it is required that $\max (x_i - x_{i-1}) \to 0$ as $n \to \infty$ then the set of limit points of (1) is the closed interval (L, U) where

$$L = \underline{\int_a^b} f(x)dx, \quad U = \overline{\int_a^b} f(x)dx.$$

This can easily be shown by taking two sequences of sums

$$\sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) \to L, \quad \sum_{i=1}^{n} f(\xi'_i)(x_i - x_{i-1}) \to U,$$

and then taking an appropriate combination of terms from each sum.

In the present note we are concerned with a generalization of these ideas. Let

$$F(x) = [f_1(x), f_2(x), \ldots, f_k(x)]$$

be a bounded vector function in the k-dimensional Euclidean space, f_j defined on $[a, b]$, $j = 1, 2, \ldots, k$. We shall consider the set of limit points as $n \to \infty$ and $\max (x_i - x_{i-1}) \to 0$ of the vector

$$(2) \qquad V_n = [S_n^1, S_n^2, \ldots, S_n^k],$$

where

$$S_n^j = \sum_{i=1}^{n} f_j(\xi_i)(x_i - x_{i-1}), \quad j = 1, 2, \ldots, k.$$

We show as a principal result that these limit points form a convex set. As we were completing this study it came to our notice that this has also been shown by P. Hartman (1947). Hartman's methods are quite different. Furthermore, our work contains incidental results which would seem to be of some interest.

THEOREM I. *Let E and E' be two everywhere dense sets on $[a, b]$ which are such that*

$$(3)\qquad \sum_{i=1}^{n} f_j(\xi_i)\,(x_i - x_{i-1}) \to L'^j,\quad \sum_{i=1}^{n} f_j(\xi'_i)\,(x_i - x_{i-1}) \to L^j,$$

ξ any point of E on $[x_{i-1}, x_i]$ ξ'_i any point of E' on $[x_{i-1}, x_i]$. Then if P is any point on the line joining the points $P_1 = (L^1, \ldots, L^j)$, $P_2 = (L'^1, \ldots, L'^j)$ there is a sequence $\Gamma_n = [S_n^1, \ldots, S_n^k]$ tending to P.

Let ρ be any real number with $0 < \rho < 1$. Let (x_{i-1}, x_i) be a subdivision of $[a, b]$, and let x_i^0 be a point on $[x_{i-1}, x_i]$ which is such that $(x_i^0 - x_{i-1}) = \rho(x_i - x_{i-1})$. Choose $\xi_i \epsilon E$, ξ_i on $(x_{i-1}, x_i^0]$. Then

$$\sum_{i=1}^{n} f_j(\xi_i)\,(x_i^0 - x_{i-1}) = \rho \sum_{i=1}^{n} f_j(\xi_i)\,(x_i - x_{i-1}),$$

and the right side tends to $\rho L'^j$ as $n \to \infty$. Let ξ'_i be any point of E' on $[x_i^0, x_i]$. Then

$$\sum_{i=1}^{n} f_j(\xi'_i)\,(v_i - x_i^0) = (1 - \rho) \sum_{i=1}^{n} f_j(\xi'_i)\,(x_i - x_{i-1}),$$

and the right side tends to $(1 - \rho)L^j$. It then follows that

$$(4)\qquad \sum_{i=1}^{n} f_j(\xi_i)\,(x_i^0 - x_{i-1}) + \sum_{i=1}^{n} f(\xi'_i)\,(x_i - x_i^0)$$

tends to $\rho L'^j + (1 - \rho)L^j = L^j + \rho(L'^j - L^j)$, and is arbitrarily near this limit for max $(x_i - x_{i-1})$ sufficiently small, $j = 1, 2, \ldots, k$.

Let E_η be the points x on $[a, b]$ at which the saltus of $f_j(x)$ over E and over E' is not less than η, $j = 1, 2, \ldots, k$. The set E_η is closed. It follows from the existence of the limits in (3) that the content of E_η is zero. Now let (x_{i-1}, x_i) be any subdivision of $[a, b]$ $[x_l, x_l',]$ the closed intervals of this set containing no points of E_η and on which the oscillation of $f_j(x)$ over E and E' is less than η, (x_k, x'_k) the remaining intervals of this set. Let max $(x_i - x_{i-1})$ be so small that the sum (3) does not differ from its limit by more than η, $j = 1, 2, \ldots, k$, and such that $\Sigma(x'_k - x_k) < \eta$, and let x_i^0 be defined as above. Let (t_{i-1}, t_i) be any subdivision of $[a, b]$, (t_p, t_p^0) the intervals of this subdivision contained on the intervals (x_l, x_l^0) of (4), (t_q^0, t_q) the intervals contained on (x_l^0, x'_l). Let ξ_p be any point of E on (t_p, t_p^0), ξ'_q be any point of E' on (t_q^0, t_q). As the maximum of $(t_i - t_{i-1}) \to 0$, $\Sigma_p(t_p^0 - t_p) \to (x_l^0 - x_l)$, $\Sigma_q(t_q - t_q^0) \to (x_l - x_l^0)$. If ξ_i is any point of

E on (x_l, x_l^0) then $|f_j(\xi_l) - f_j(\xi_p)|$ is not greater than η, and if ξ'_l is any point of E' on (x_l^0, x'_l) then $|f_j(\xi'_l) - f(\xi'_q)|$ is not greater than η. Hence for max $(t_i - t_{i-1})$ sufficiently small

$$\sum_p f_j(\xi_p)\,(t_p^0 - t_p) + \sum_q f(\xi'_q)\,(t_q - t_q)$$

does not differ from $f_j(\xi_l)\,(x_l^0 - x_l) + f(\xi'_l)(x_l - x_l^0)$ by more than $\eta(x_l - x_{l-1}) + \eta_l$ where $\Sigma\eta_l < \eta$. Hence if ξ_l is any point of (t_l, t'_l) for the intervals (t_k, t'_k) of the set (t_{i-1}, t_i) not included in the intervals (t_p, t_p^0) and (t_q^0, t_q) for all the intervals (x_{l-1}, x_l), the sum

$$\sum f_j(\xi_l)\,(t'_k - t_k) + \sum f_j(\xi_p)\,(t^0_p - t_p) + \sum f_j(\xi'_q)\,(t_q - t_q^0)$$

does not differ from the sum (4) by more than $\eta M + \eta(b - a) + \eta$ where $|f_j(x)| < M$, $j = 1, 2, \ldots, k$, it being understood that the second and third sums are taken for all the intervals (x_l, x'_l). The number η is arbitrary. Hence we conclude that if (x_{i-1}, x_i) is an arbitrary subdivision of $[a, b]$ there is a choice of Σ''_i on (x_{i-1}, x_i) for which

$$\sum f(\xi''_i)\,(x_i - x_{i-1}) \rightarrow L^j + \rho(U^j - L^j)$$

as max $(x_i - x_{i-1}) \rightarrow 0$, $j = 1, 2, \ldots, k$. If now P is any point on the line joining P_1 and P_2 ρ can be so chosen that

$$[L' + \rho(U' - L'), \ldots, L^k + \rho(U^k - L^k)]$$

is the point P. Hence the theorem.

If $|(f_j(x)| < M$, then there is a choice of ξ_i and of ξ'_i on (x_{i-1}, x_i) for which

$$\sum f_j(\xi_i)(x_i - x_{i-1}) \rightarrow U^j = \int_a^b f_j(x)dx,\quad \sum f_j(\xi'_i)\,(x_i - x_{i-1}) \rightarrow L^j = \int_a^b f_j(x)dx$$

However, we have not succeeded in showing that there are everywhere dense sets of E and E' such that ξ_i can be any point of E and ξ'_i any point of E'. In order to determine the distribution of the limit points of (2) for the general case we make use of the following well-known results of the real variable theory.

Let C be a closed set on $[a, b]$ and let $f(x)$ be continuous over C. Let (x_{i-1}, x_i) be a sequence of sub-divisions of $[a, b]$ for which max $(x_i - x_{i-1})$ tends to zero. Let (x_k, x'_k) be the intervals of this sub-division which have points of C on their interior. Then if ξ_k is any point of C on (x_k, x'_k) the sum $\Sigma f(\xi_k)(x'_k - x_k)$ tends to a limit which we shall call the Riemann integral of $f(x)$ over C. The proof of the existence of this limit does not differ in any essential way from the proof of the existence of the ordinary Riemann integral of a function continuous on $[a, b]$. It can also be easily shown that this integral over C is equal to the Lebesgue integral of $f(x)$ over C,

$$R\int_C f(x)dx = L\int_C f(x)dx.$$

Now let $f(x)$ be bounded on $[a, b]$ and measurable. Since $f(x)$ is measurable there is a closed set C on $[a, b]$ with $|C|$, the measure of C, satisfying $b - a - |C| < \epsilon$ where ϵ is arbitrary, with $f(x)$ continuous on C, and with

$$\left| L\int_a^b f(x)dx - L\int_C f(x)dx \right| < \epsilon.$$

Let the functions $f_j(x)$ be such that $|f_j| < M, j = 1, 2, \ldots, k$. Let (x_{i-1}, x_i) be a subdivision of $[a, b]$. Suppose there is ξ_i on (x_{i-1}, x_i) for which

$$[S_n^1, S_n^2, \ldots, S_n^k] \rightarrow [U^1, U^2, \ldots, U^k],$$

and a point ξ'_i for which

$$[S_n^1, S_n^2, \ldots, S_n^k] \rightarrow [L^1, L^2, \ldots, L_n^k].$$

Let $\psi_n^j(x) = f_j(\xi_i)$ on $x_{i-1} \leqq x < x_i$, $i = 1, 2, \ldots, n-1$, $\psi_n^j(x) = f_j(\xi_n)$ on $x_{n-1} \leqq x \leqq x_n$. ψ_n^j is a step-function and consequently summable. Furthermore,

$$\int_a^b \psi_n^j dx \rightarrow U^j.$$

Hence, by the Mean Convergence Theorem there is a function $\psi^j(x)$ for which

$$\int^b |\psi_n^j - \psi^j| \, dx \rightarrow 0$$

and a subsequence $\psi_{n_\nu}^j(x)$ which tends to $\psi^j(x)$ for almost all x on $[a, b]$. This subsequence $\psi_{n_\nu}^j(x)$ evidently can be so chosen that the convergence is independent of j. It is then possible to find a closed set C on $[a, b]$ with $b - a - |C| < \epsilon$, with ψ^j continuous on $C, j = 1, 2, \ldots, k$ and with $\psi_{n_\nu}^j \rightarrow \psi^j(x)$ uniformly on $C, j = 1, 2, \ldots, k$.

Let (x_{i-1}, x_i) be any subdivision of $[a, b]$, and let (x_l, x'_l) be the intervals of this set for which

$$(5) \qquad |C(x_l, x'_l)| > (1 - \epsilon)(x'_l - x_l).$$

If max $(x_i - x_{i-1})$ is sufficiently small then

(i) $\sum (x'_l - x_l) > b - a - 2\epsilon$,

(ii) $\left| \sum \psi^j(\xi_l)(x'_l - x_l) - \int_C \psi^j dx \right| < \epsilon$

where ξ_l is any point of C on (x_l, x'_l), and

(iii) $|\psi^j(x) - \psi^j(x')| < \epsilon$

where x and x' are any two points of C on (x_l, x'_l) $j = 1, 2, \ldots, k$. (iii) follows from the fact that $\psi^j(x)$, being continuous on C, is uniformly continuous on C.

Now let ρ be any number with $0 < \rho < 1$. Fix ν_0 so that

$$(6) \qquad |\psi^j_{n_\nu}(x) - \psi^j(x)| < \epsilon, j = 1, 2, \ldots, k, \nu > \nu_0,$$

for x on C. Let (t_{lp}, t_{lp}^0) be the intervals of constancy of the function $\psi_{n_\nu}^j$ on (x_l, x_l^0) where $x_l^0 - x_l = \rho(x'_l - x_l)$, $\nu > \nu_0$, and where (t_{lp}, t_{lp}^0) contains points of C. Let ξ_p be any point of C on (t_{lp}, t_{lp}^0), and let $\bar{\xi}_p$ be the point of this interval for which $\psi_{n_\nu}^j = f_j(\bar{\xi}_p)$. Then (6) is satisfied for $x = \xi_p$ and $\psi_{n_\nu}^j(\xi_p) = f_j(\bar{\xi}_p)$. Consequently

$$|\psi^j(\xi_p) - f_p(\bar{\xi}_j)| < \epsilon,$$

and if ξ_l is any point of C on (x_l, x'_l) it follows from (iii) that

$$|\psi^j(\xi_l) - f(\bar{\xi}_p)| < 2\epsilon,$$

this relationship holding for every interval (t_{lp}, t_{lp}^0) on (x_l, x_l^0), ξ_l any point of C on (x_l, x'_l), $j = 1, 2, \ldots, k$. Then

$$\sum_p f(\bar{\xi}_p)(t_{lp}^0 - t_{lp}) = [\psi^j(\xi_l) + \epsilon_l] \sum_p (t_{lp}^0 - t_{lp})$$

where $\epsilon_l < 2\epsilon$. From (5) it follows that

$$\sum_p (t_{lp}^0 - t_{lp}) = x_l^0 - x_l - \eta'_l$$

where $0 < \eta' < \epsilon(x'_l - x_l)$. We then have

$$\begin{aligned} \sum_p f_j(\bar{\xi}_p)(t_{lp}^0 - t_{lp}) &= [\psi^j(\xi_l) + \epsilon_l](x_l^0 - x_l - \eta'_l) \\ &= [\psi^j(\xi_l) + \epsilon_l][\rho(x'_l - x_l) - \eta'_l]. \end{aligned}$$

Summing over l gives on the right side

$$\rho \sum_l \psi^j(\xi_l)(x'_l - x_l) - \sum_l \eta'_l \psi^j(\xi_l) + \rho \sum_l \epsilon_l (x'_l - x_l) - \sum_l \epsilon_l \eta'_l.$$

It then follows from (ii) and the restrictions on ϵ_l and η'_l that

$$\sum_l \sum_p f_j(\bar{\xi}_p)[t_{lp}^0 - t_{lp}) \text{ and } \rho \int_C \psi^j(x)dx$$

do not differ by more than $\epsilon M(b - a) + \rho\epsilon(b - a) + \epsilon^2(b - a)$, where $|\psi^j| < M$, $0 < \rho < 1$, and ϵ is arbitrary, $j = 1, 2, \ldots, k$. We note that the difference between

$$\rho \int_C \psi^j(x)dx \text{ and } \rho \int_a^b \psi^j(x)dx = \rho L^j$$

can be made arbitrarily small by taking C with $|C|$ sufficiently close to $b - a$.

Setting $\varphi_n^j(x) = f_j(\xi'_i)$ on $x_{i-1} \leqq x < x_i$, $i < n$, $\varphi_n^j(x) = f_j(\xi'_n)$ on $x_{n-1} \leqq x \leqq x_n$ and proceeding as above, we arrive at a set of intervals (t_{lq}^0, t_{lq}) belonging to the subdivision (t_{i-1}, t_i) of (a, b), with (t_{lq}^0, t_{lq}) on (x_l, x'_l) and with

$$\sum_l \sum_q f_j(\xi_q)(t_{lq} - t_{lq}^0), j = 1, 2, \ldots, k,$$

arbitrarily near to $(1 - \rho)L^j$ provided the subdivision (t_{i-1}, t_i) of $[a, b]$ has max $(t_i - t_{i-1})$ sufficiently small. Now let (t_m, t'_m) be the intervals of the set (t_{i-1}, t_i) which do not belong to either of the sets (t_{lp}, t_{lp}^0) or (t_{lq}^0, t_{lq}^0), and let ξ_m be any point on $(t_m t'_m)$. Then

$$\sum |f_j(\xi_m)(t'_m - t_m)| < M \sum (t'_m - t_m)$$

which is arbitrarily small if the sum of the lengths of the intervals

(t_{lp}, t_{lp}'), (t_{lq}^0, t_{lq}) is sufficiently close to $b - a$. We can now conclude that there are sums

$$\sum f_j(\xi_i)(x_i - x_{i-1}) = \sum_m f_j(\xi_m)(t'_m - t_m) + \sum_l \sum_p f_j(\bar{\xi}_p)(t_{lp}^0 - t_{lp}) \\ + \sum_l \sum_q f_j(\bar{\xi}_q)(t_{lq} - t_{lq}^0)$$

which are arbitrarily near to

$$L^j + \rho(U^j - L^j),\ j = 1, 2, \ldots, k.$$

Hence this point is a limit point of $V_n = [S_n^1, S_n^2, \ldots, S_n^k]$. Since ρ is any number with $0 < \rho < 1$ it follows that every point on the line joining

$$P_1 = [U^1, U^2, \ldots, U^k] \text{ and } P_2 = [L^1, L^2, \ldots, L^k]$$

is a limit point of V_n. The points P_1 and P_2 are any two points in the set A of limit points of V_n. Hence we have established

THEOREM II. *Let $F(x)$ be a vector function,*

$$F(x) = [f_1(x), f_2(x), \ldots, f_k(x)]$$

$a \leqq x \leqq b$ in k dimensional Euclidean space. Let $|f_j| < M$, $j = 1, 2, \ldots, k$. Let (x_{i-1}, x_i) be a subdivision of $[a, b]$, ξ_i any point on $[x_{i-1}, x_i]$. Then the set A of limit points of

$$\left[\sum_{i=1}^{n} f_j(\xi_i)(x_i - x_{i-1}), \ldots, \sum_{i=1}^{n} f_k(\xi_i)(x_i - x_{i-1})\right]$$

as max $(x_i - x_{i-1}) \rightarrow 0$ *is a closed convex set.*

If all the functions f_j are Riemann integrable A is a single point. If $f_1 = f_2 = \ldots = f_k$, A is a linear segment. If all the functions are Riemann integrable save one, A is a linear segment. In general the set A can be of any dimension less than or equal to k depending on the nature of the functions $f_1, \ldots, f_j$.

The following theorems on the Lebesgue integral of a function are a consequence of the foregoing proofs.

THEOREM III. *Let $f_j(x)$, $j = 1, 2, \ldots, k$, be measurable and bounded on (a, b). Given $\epsilon > 0$ there exists a set E on (a, b) such that if (x_{i-1}, x_i) is any subdivision of (a, b) with* max $(x_i - x_{i-1})$ *sufficiently small then*

$$\left|\sum f_j(\xi_i) - \int_a^b f_j(x)dx\right| < \epsilon,\ j = 1, 2, \ldots, k,$$

where ξ_i is any point of E on (x_{i-1}, x_i) provided there are points of E on this interval. Otherwise ξ_i is any point on (x_{i-1}, x_i).

REFERENCES

HARTMAN, P. (1947). Limits of Riemann approximating sums. Quar. Jour. of Math., **18**: 125-7.

SASTRY, R S (1940). The limiting points of the set of the upper and lower sums in Riemann integration. The Mathematics Student, no. 8: 148-50.

26 MAR 1951

TRANSACTIONS OF

THE ROYAL SOCIETY OF CANADA

SECTION III

CHEMICAL, MATHEMATICAL,
AND
PHYSICAL SCIENCES

THIRD SERIES—VOLUME XLIV—SECTION III

JUNE, 1950

56672

OTTAWA
THE ROYAL SOCIETY OF CANADA
1950

TABLE OF CONTENTS

The Atmospheres of the Early Type Stars. Presidential Address. By C. S. Beals, F.R.S.C. 1

The Action of Hydrogen Peroxide—Nitric Acid Mixtures on Ketones. By G. H. Bjorklund and W. H. Hatcher, F.R.S.C. 25

The Steinitz-Gross Theorem on Sums of Vectors. By Ira Damsteeg and Israel Halperin 31

Examples of Integrals that are Discontinuous in Sets of Positive Measure. By H. W. Ellis 37

Limit Points of Riemann Sums. By R. L. Jeffery, F.R.S.C. . 43

Correlation of Auroral Observations in the Northern and Southern Hemispheres. By D. E. Little and G. M. Shrum, F.R.S.C. 51

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLIV : SERIES III : JUNE, 1950
SECTION THREE

PRESIDENTIAL ADDRESS

The Atmospheres of the Early Type Stars*

By C. S BEALS, F.R S C.

THOSE of us who were responsible for planning this symposium almost a year ago probably did not fully realize what a large part of astrophysics our subject took in. A glance over some of the major publications devoted to the physics of the stars during the past two decades reveals the fact that well over half the titles deal either directly or indirectly with the atmospheres of the stars and the planets. Under such circumstances, it is scarcely necessary to say that we cannot undertake anything like a complete summary of this field. Instead, the papers presented will stress a few major developments in this part of astrophysics and where possible endeavour to link them with interesting aspects of terrestrial physics. In so far as my own contribution is concerned it has served to emphasize again something that scarcely needs emphasis, namely, that physics and astrophysics are one subject and that there is no dividing line between them.

While this paper deals with the atmospheres of the early type stars, it is scarcely possible to discuss stellar atmospheres without making at least some reference to the structure of a star as a whole. It is very pleasant, in fact, to have an occasion to refer to some of the results which have been achieved in the study of stellar interiors since this is one of the brightest chapters in the whole history of modern science. While the interior of a star is not susceptible to direct observation, the application of the methods of mathematical physics to gas spheres of appropriate mass and chemical composition by Lane [1], Emden [2], Eddington [3], Jeans [4], Chandrasekhar [5], and a host of others has led to a reasonably logical and consistent picture of both the inside and the outside of a star. As a result of these studies combined with the results of observational astrophysics, we may now speak with some confidence of the following broad facts concerning stars.

(1) The stars range in diameter from around 500,000,000 km. to about 20,000 km.

(2) The masses of the stars show a less drastic variation from the mean. The most massive stars are from 100 to 200 times the sun while the so-called white dwarfs are from 1/5 to 1/10 the solar mass.

---

*Part of a symposium on the Atmospheres of the Stars and the Planets.

(3) While the densities of the outer layers are of the order of 10^{-4} atmosphere or less, the density of the deep interior is always quite large. A good mean value of the central density would probably be of the order of 100 g. per c.c. but for the white dwarfs it may be as great as 10^6 g. per c.c.

(4) In spite of these great densities, the temperatures in the deep interiors of stars are so high (of the order of 30,000,000°K.) that the stars are believed to be gaseous throughout their volumes right to the centre. The term gaseous means that the material of the stellar volume behaves approximately as a perfect gas and, as you are well aware, this is possible because the atoms are highly ionized so that the electrons and the atomic nuclei together constitute the molecular units of the gas.

(5) The central temperatures which, as already mentioned, are of the general order of 20 to 30 million degrees, not only make possible the existence in a gaseous state of matter at great densities but the pressure of radiation is so great as to be an important factor in maintaining equilibrium and preventing the star from collapsing by its own weight.

From the point of view of the present discussion, the most important fact of this brief summary is that stars are completely gaseous bodies with no such sharp distinction between atmosphere and inner surface as is found, say, on the earth or the other planets. In a certain sense it might be correct to say that a star is all atmosphere, although it is necessary to distinguish between the different layers in any quantitative discussion of observed phenomena. For purposes of clarity, we will, therefore, arbitrarily restrict the atmosphere to that part of the star which is susceptible to direct observation by optical methods.

Consider the optical thickness τ of a stellar atmosphere as defined by the equation

$$I_\lambda = I_{0\lambda} 10^{-\tau_\lambda}. \tag{1}$$

Here I_0 is the intensity of stellar radiation at optical depth τ, and I_λ is the intensity reaching the observer. It will be noted that increasing values of τ lead to smaller and smaller values of I_λ though there is theoretically no limit beyond which observations cannot penetrate. Most actual astrophysical observations, however, will be concerned with optical depths less than 3, and for most practical purposes the depth of the stellar atmosphere may be defined in this way. Since τ is a function of λ the linear extent of the observable atmosphere will depend upon the wave-length of the radiation involved.

The actual linear depth of a stellar atmosphere is thus never known with any degree of certainty though we are accustomed to think in terms of depths of thousands or even hundreds of thousands of kilometres. Of one point, however, we may be reasonably certain; for an ordinary absorption line star the depth of the observable stellar atmosphere is an insignificant fraction of the radius of the star as a whole.

Classification. Since this contribution is restricted to the hotter stars it seems well to make a brief reference to the subject of classification. The Harvard system divides the stars into the classes W, O, B, A, F, G, K, M, N, R, S, representing, as we shall see later, a temperature scale ranging from approximately 100,000°K. to 2,000°K. Roughly speaking the hotter stars may be placed in the range having temperatures greater than 10,000°. This will include the early A's and all stars of earlier spectral class.

Stark effects in stellar spectra. One of the most interesting phenomena of stellar spectroscopy is the tremendous width of hydrogen lines in certain A type stars, especially when these widths are contrasted with narrow metallic lines observed in the same spectra. Fig. 1 shows the line Hγ of α^2 Geminorum which has a width of 200 angstroms, while clearly visible on the same spectrum are numerous metallic lines which are relatively fine and narrow. Another type of observation of almost equal interest is found in the fact that not all hydrogen lines of A type stars exhibit the phenomenon of broad lines. Fig. 1 also shows profiles of Hγ for θ Aurigae and H.D. 223385 as well as α^2 Geminorum. These three stars are all of early A type and yet the line widths show a progressive decrease in passing from α^2 Gem. to H.D. 223385. R. M. Petrie [6] has discussed this phenomenon and has shown that the hydrogen line width is closely correlated with the intrinsic luminosity of the stars, wide lines being associated with stars of low luminosity and vice versa. The very luminous star H.D. 223385 is approximately 900 times as bright as α^2 Geminorum and 175 times as bright as θ Aurigae.

There is now little ambiguity attaching to the nature of the wide hydrogen lines of A type stars. A logical explanation is found in Stark effect due to the electric fields which are present during the close approaches of radiating atoms to charged ions in a highly ionized stellar atmosphere.

The dependence of the absorption coefficient on the temperature and electron pressure is illustrated in Equation 2 due to Unsold [7] and Holtzmark [8].

$$- \quad - \quad \frac{CF_0^{3\ 2}}{\Delta\lambda^{5\ 2}} \tag{2}$$

Here F_0, the average field, has the following dependence on the electron pressure

$$F_0 = 46.8\left(\frac{2P_e}{T}\right)^{2\ 3}. \tag{3}$$

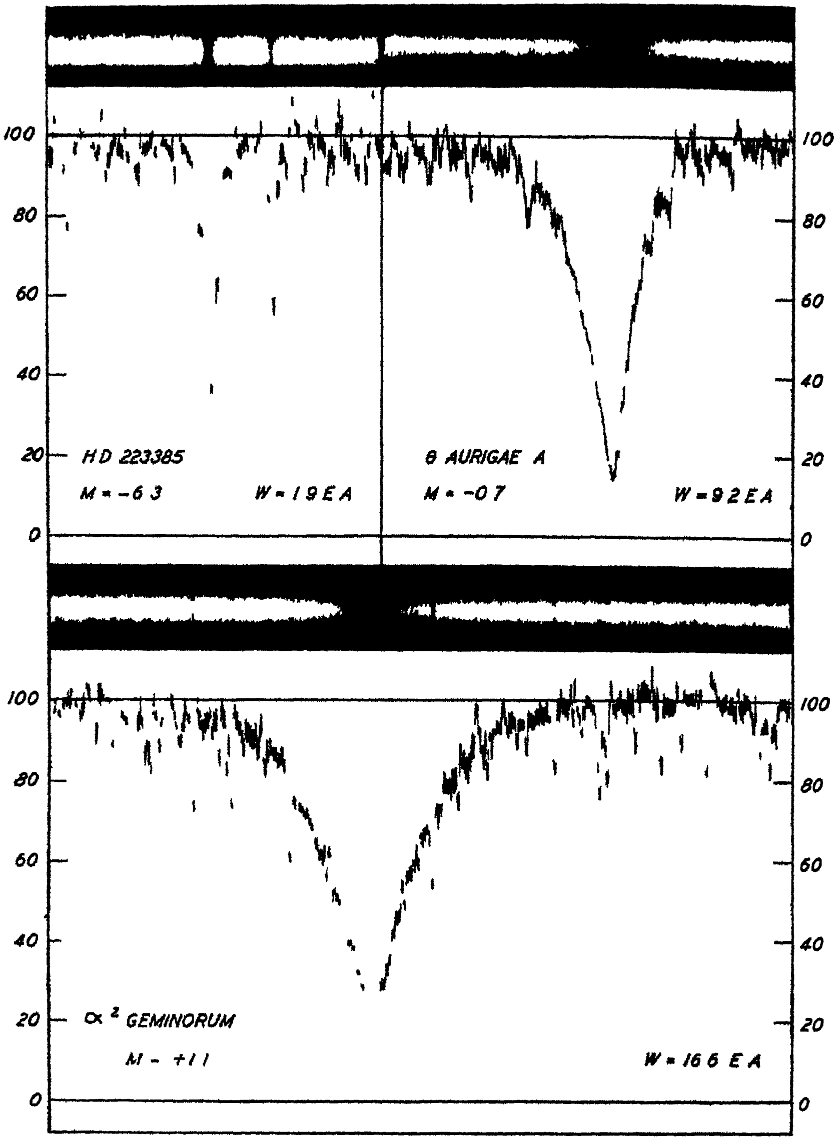


FIG 1. Stark effect in A type stars Spectra and intensity records of the region of $H\gamma$ in three typical A type stars The remarkable increase in Hydrogen absorption, from the supergiant H D 223385 to the dwarf α^2 Geminorum, is due to the extensive line-wings typical of the Stark effect The $H\gamma$ line in α^2 Geminorum is definitely stronger than that in θ Aurigae as shown by the intensity records This difference is difficult to appreciate from a visual inspection of the spectrograms. (Figure kindness of Dr R M Petrie)

Here $\Delta\lambda$ is the wave length measured from the centre of the observed hydrogen line, P_e is the electron pressure, and T is the temperature.

The observed correlation between line width and absolute magnitude is thus readily explicable in terms of Stark effect. In a dwarf star of low absolute magnitude, the value of surface gravity is relatively high. The gas pressure and electron pressure are correspondingly high providing conditions favourable for the close approach of atoms and ions during radiation and absorption processes. For a star of large absolute magnitude, the reverse is the case, the density and electron pressure are lower and the average field strength is smaller due to less frequent encounters of atoms and ions.

Of the atoms other than hydrogen, the only one likely to exhibit Stark effect of observable magnitude is helium. The line displacements, however, are in general only half those expected for hydrogen and since the abundance of neutral helium in the B type stars is very much smaller than that of hydrogen in the early A's, the effects are more difficult to observe. Three general types of effects may be expected

(1) For stars where the lines are sufficiently narrow, it may be expected that Stark effect will widen or distort the profile in a more or less predictable manner, usually in an unsymmetrical way.

(2) The existence of electric field may be expected to result in the appearance of certain forbidden transitions known as electric combination lines, the position of which may be predicted from atomic theory. They should in general be immediately to the violet of the normal line position.

(3) For stars where the detailed line profiles are obscured by rotational broadening, the Stark effect may show itself by an anomalous distribution of intensity in the various helium series since the effect of electric fields is not the same for lines of different series numbers.

Since for a number of reasons it is more difficult to draw general conclusions from (3), our present discussion will be confined to (1) and (2).

In actual observations (1) and (2) ordinarily act together, the Stark effect on the normal line producing a line unsymmetrically extended to the red, while the electric combination lines which appear immediately to the violet of the normal line are unsymmetrically extended to the violet. The net result may be a stellar line which approaches symmetry quite closely and which suggests a miniature of the broadened hydrogen profiles. For a few stars with lines sufficiently sharp, however, what appear to be the electric combination lines are present as definitely observable minima to the violet of the normal line position. This was at first hailed [9] as indisputable evidence of the presence of helium Stark effect in stellar spectra; but later studies by Foster and Douglas [10] and by Underhill and Petrie [11] showing that their minima did not appear in the predicted position have introduced a considerable element of uncertainty. These effects are clearly shown in the above-mentioned paper by Foster and Douglas and in a more recent article describing the helium line profiles in B type stars by Underhill [12]. The present position with regard to helium Stark effect in stellar atmosphere is, therefore, not entirely satisfactory. Few astrophysicists would doubt its existence, but the observed phenomena have not been completely explained and it is possible that further astrophysical studies of this phenomenon will lead to new knowledge concerning the effects of electric fields on radiating or absorbing atoms.

Motions in stellar atmospheres. Apart from temperature agitations common to all assemblages of atoms or molecules there are several types of motions in the atmospheres of stars which are revealed by studies of stellar spectra. The first of these is rotation of the star as a whole. The existence of rotation among bodies of stellar and planetary character seems well nigh universal and has numerous examples in the solar system. The existence of spectroscopic and visual binary stars is another clear indication of rotation among celestial bodies. Further, a close examination of the phenomena of eclipsing binaries shows clearly that in some instances the rotation of the component stars is more rapid than that indicated by the period of the binary system as a whole.

Quite apart from studies of binaries there is very good evidence from the spectra of individual objects that most early type stars are in a state of rapid rotation. This is indicated by a broadening and flattening of the stellar line profiles which is more marked as the rotation increases.

The velocity in the line of sight of any point on a stellar disc is given by the formula:

$$V = \omega r_1 \sin i \tag{4}$$

where ω is the angular velocity of rotation, r_1 is the projection on the disc of the distance from the point considered to the axis of rotation and i is the inclination of the axis of rotation to the line of sight.

The effect which rotation of the star will have on an absorption line may be found by integrating over the disc on the basis of an assumed initial form of the line. Such calculations have been carried out by Struve [13] and Carroll [14], who have shown that the broadened forms of many stellar lines are due to rotation. The velocities in the line of sight due to rotation may range from a few kilometres per second to several hundred km./sec. depending on the velocity of rotation and the inclination of the axis of rotation.

The rotational broadening approaches zero for small values of the angle i and it is generally supposed that early type stars with sharp lines have their axes of rotation closely coinciding with the line of sight. On this assumption the proportion of narrow lined stars is readily calculable and the observations are closely in agreement with prediction. For the most part, relatively little attention has been paid to rotation as a major influence on the physics of the stellar atmosphere, but there may well be some instances where the centrifugal force of rotation is sufficient to upset normal equilibrium and cause the streaming out of matter from the equatorial plane. Various phenomena of stellar emission have been attributed to this cause by Struve and it may well be that it is in part responsible for the existence of the large group of B type emission line stars.

Another type of motion in stellar atmospheres which is known to exist but which depends on somewhat less clear-cut observational evidence is turbulence. Turbulence, as distinguished from temperature motions of individual atoms, may be defined as the motion of groups of atoms which may in some instances constitute quite large volumes of the stellar gas. As might be expected, line profiles due to turbulence have an exponential form of the general character of equation (5), due to Struve [15].

$$\sigma = \text{const. } e^{-\left(\frac{v}{v_0}\right)^2}, \tag{5}$$

where σ is the absorption coefficient corresponding to a particular velocity displacement v in the line, and v_0 is the mean velocity due to turbulent or temperature motions of the gas molecules.

It is somewhat difficult to distinguish effects due to turbulence from those caused by rotation. They both lead to a broadened and flattened character to the centre of the line and, although the wings are different, the distinction may be observationally none too clear especially where both effects may be present in a single star.

For various late type stars the presence of turbulence has been inferred from studies of the curve of growth, the Doppler portion of the curve being extended beyond that which could be due to temperature motions alone. So far such methods have been less used for early type objects, and so present ideas on turbulence in these stars are based on such evidence as is available from line profiles and on the conviction that high surface temperatures provide favourable conditions for the development of turbulent motions.

Both stellar rotation and turbulence are a nuisance to the stellar spectroscopist who is trying to carry out a detailed analysis of a stellar atmosphere. They cause widening of the lines and make it difficult to detect the details of structure due, say, to electric or magnetic fields. This is particularly the case for studies in the very interesting field of stellar magnetism [16] and the number of early type stars with lines sufficiently sharp for studies of this kind is very few. Fortunately, the almost universal existence of rapid rotation in early type stars makes it possible to be reasonably sure of the orientation of the axis of rotation for those objects which do have fine lines.

The type of stellar atmospheric pulsation affecting the volume of he atmosphere as a whole and known as Cepheid variation [17] is well known to all, and it is not proposed to discuss it in any detail in this review. This type of atmospheric motion is mainly characteristic of the later type stars. The periods of such pulsations vary from half a day to 100 days, and the magnitude of the velocity variations is usually of the order of 120 km./per sec. There are a few early type stars which exhibit this type of motion, the best example being H.D. 199140.* In this star the magnitudes of the velocities involved are ± 60 km./sec. which in anything but a star would be considered to be a phenomenon of explosive violence.

In addition to rotation, turbulence, and pulsation, another very important type of motion in stellar atmospheres is the actual large scale ejection of material from the stellar surface into outside space. Even though the evidence is for the most part indirect, most students of the sun and its effects on earthly conditions now believe that at times there is fairly large scale ejection of atoms from the solar surface. The agency involved in the main is believed to be radiation pressure augmented by the existence of hot spots on the sun's surface. Since this phenomenon is reasonably well established for the sun, it would appear logical to suppose that the high surface temperatures of early type stars would give rise to effects similar in kind but of a more violent character.

The first definite spectroscopic evidence for the ejection of gases from a stellar surface was found for the novae [18] where it has been established that the tremendous increase in brightness is accompanied by the throwing off of gaseous layers from the surface with velocities which are sometimes as great as three or four thousand kilometres per second. Evidence for these velocities is found in the positions of the absorption lines of the Balmer series which are sometimes displaced to the violet by as much as 60 angstroms corresponding to a Doppler

---

*Unpublished work of R. M. Petrie.

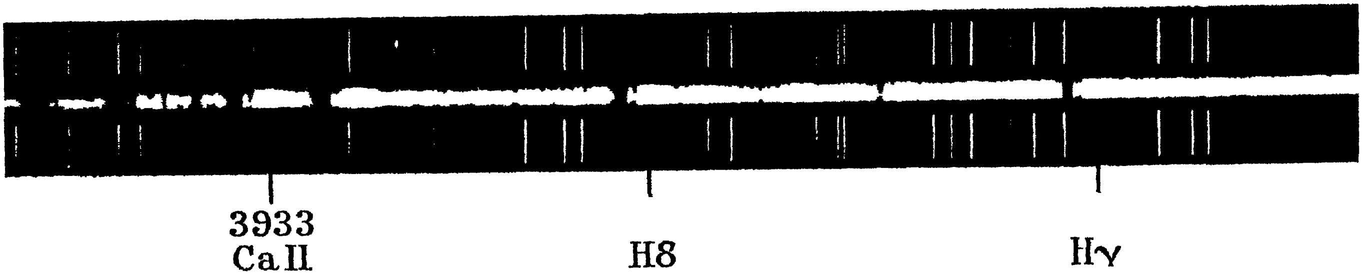


FIG. 2. The spectrum of Nova Herculis, Feb. 3, 1935. The broad emission bands with absorption borders on their violet edges are believed to be due to the ejection of material from the stellar surface. This spectrum represents a relatively early stage in the nova outburst when the temperature is still comparatively low

effect of 4,000 km./sec. The effect is illustrated in Fig. 2 showing the spectrum of Nova Herculis on February 3, 1935.

It will be noted that there are no corresponding absorption lines to the red and that numerous atoms in the spectrum in addition to the Balmer series are similarly affected so that the interpretation of the observations is scarcely in doubt. To make the matter more certain actual increases in the observed diameter of the star in agreement with the velocities derived from the spectra have been observed for certain novae whose distances are known.

Spectroscopic evidence of an essentially similar kind indicates that large scale ejection of material takes place from the atmospheres of the Wolf Rayet [19] and P Cygni Stars [20] also. For the Wolf Rayet Stars, the displacement to the violet of the absorption lines corresponds in general to velocities from 1,000 to 4,000 km./sec. For the P Cygni stars the ejection velocities are generally a few hundred km./sec. These displaced absorption lines are commonly interpreted as due to the ejection of gases and the bright lines which accompany them are believed to arise in the envelope of ejected atoms which surrounds the star. For most early type stars which exhibit this phenomenon the velocity of ejection exceeds the parabolic velocity of escape so that the material is eventually lost to the star.

Although only for the types of stars mentioned above is the fact of ejection established by direct spectroscopic evidence, it is probable that it is a feature of all early type stars. While, as has already been suggested, the process may be assisted in some instances by the centrifugal force of rapid rotation, the major accelerating force probably has its origin in the high temperature of the stellar surface which acts either through direct radiation pressure or in some other manner not now fully understood.

TEMPERATURES OF ABSORPTION LINE STARS

The continuous spectrum. So many of the characteristics of a stellar atmosphere are dependent on its temperature that determinations of this quantity by direct or indirect means are of major importance in this branch of astrophysics. The initial approach to the problem of stellar temperature is to regard the star as a black body which radiates according to Planck's formula as expressed in the equation:

$$E_\lambda = \frac{C_1\lambda^{-5}}{\epsilon^{\frac{C_2}{\lambda T}} - 1}, \tag{6}$$

where $C_1 = 3.71 \times 10^{-5}$, $C_2 = 1.435$, λ expressed in centimetres, ϵ = base of nat. logs., and E_λ = radiant energy in ergs per cm.^{2} per sec. in the wave-length range.

Making use of this formula it is possible to derive Wien's law for the wave-length of radiation of maximum intensity

$$\lambda_m = \frac{0.289}{T}\text{cm.}, \tag{7}$$

and also Stefan's law for the total quantity of radiation in all wave-lengths emitted by a star of given diameter and temperature

$$E = 5.72 \times 10^{-5}T^4 \tag{8}$$

where E = ergs per cm.^{2} per sec. While theoretically either of these two quantities can be used for the determination of temperature, actually for the early type stars most of the radiation is in the far ultraviolet which cannot be directly observed. As a consequence, Wien's law becomes impossible of application.

It is possible to make an estimate of the total radiation corresponding to the so-called effective temperature but the fact that most of the radiation is in a region inaccessible to observation makes this difficult, and the matter is further complicated by uncertainty in the distances and intrinsic luminosities which are necessary for this method of estimating temperature.

In general, therefore, the application of Planck's law to studies of the continuous spectrum is limited to comparisons of the slope, expressed in suitable units, of the energy curve in the observable region of the spectrum with that of a body of known temperature. For relative values, standard stars may be used but for absolute determinations of temperature a terrestrial source is necessary. Unfortunately there are a number of difficulties in applying these methods to stars and they may be listed as follows:

(1) Owing to the presence of absorption lines and of absorption at the limits of optical spectral series such as the Balmer series of hydrogen, it is known that stars show deviations from black body radiation, and these deviations from black body laws may introduce considerable errors into temperature determinations.

(2) For early type stars with temperatures in excess of 20,000°K. the change of slope of the energy curve with temperature is so small as to make the method difficult of application.

(3) In comparisons with a terrestrial source and to a lesser extent for inter-comparisons of stars the variable absorption of the earth's atmosphere is difficult to allow for with precision. Atmospheric ab-

sorption not only varies with altitude, but it also varies from night to night and even from moment to moment, introducing an unavoidable source of uncertainty into gradient measurements.

(4) Finally for the early type stars, which are mostly fairly distant, a major disturbing factor is found in the differential absorption of light by interstellar matter. General absorption by interstellar material is believed to be due to small particles of the order of 10^5 cm. diameter and it has been shown that the law of absorption varies with wavelength according to a law which is expressed in the following equation:

$$\sigma = \text{const.} \times \lambda^{-1}. \tag{9}$$

W. Petrie [21] has recently made an attempt to derive stellar gradients for a number of early type stars by eliminating the absorption of interstellar matter on the basis of distances derived from interstellar line intensities. He concluded that insterstellar material, whether gas or dust, is too irregularly distributed for such a method to be successful and that only nearby stars can be dealt with in this way.

It will be understood of course that in spite of the above-mentioned difficulties there have been many determinations of temperature by the observation of gradients and a temperature scale has been worked out which indicates temperatures of 10,000° to 14,000° for the early A stars to 30,000° for the hotter O's. It is questionable, however, on the basis of the direct application of Planck's law to the observable continuous spectrum whether we know the temperature of the AO stars, which form the reference point for the early type stars as a group to an accuracy of better than 3,000° to 5,000°. It is, therefore, necessary to turn to other methods if we wish to have values of temperature in which we can place full confidence.

Ionization in stellar atmospheres. A very superficial examination of the spectra of the early type stars will reveal the presence of absorption lines due to atoms of high ionization potential such as HeI, HeII, OII, OIII, NII, NIII, NIV, NV, CII, CIII, CIV and many others. The general type of relationship which exists between temperature, density, and state of ionization in a normal star is expressed in the following equation due to Saha [22] :

$$\text{Log}\,\frac{N_1}{N_0} = -\frac{5040}{T}\,(\text{I.P.}) + 2.5 \log T - 0.48 + \log\frac{2B_1}{B_0} - \log P_e. \tag{10}$$

N_0 and N_1 are number of atoms in neutral and ionized states respectively. B is the partition function, while P_e is the electron pressure which depends largely on the density of the atmosphere. I.P. and T are of course ionization potential and temperature.

It is of some interest to point out that formulae of this general character were first used in studies of the state of ionization of chemical solutions. The problem of using formulae of this type to deduce the temperatures of stellar atmospheres is quite an old one. In the early studies by Saha [22], Fowler and Milne [23], Payne [24], and others, only eye estimation of line intensities were available and efforts were directed towards finding a suitable relationship between the temperature of the star and the appearance of a certain atom or ion either at maximum intensity or marginal appearance. Both of these points were difficult to determine accurately and the latter especially was very dependent on the quality and dispersion of the available spectra.

Many of these difficulties have been avoided in a recent investigation by R. M. Petrie [25], who has utilized precise spectrophotometric measures of the intensities of lines of HeI and HeII as his observational data. He has shown that by suitably combining Boltzmann's formula for the distribution of atoms in excited states

$$\frac{N_a}{N_0} = \frac{w}{B_0(T)}\, e^{-\frac{\chi}{kT}} \tag{11}$$

(N_0, total number of atoms; N_a, atoms in excited state; χ, excitation potential; B_0, partition function; w, statistical weight) with the previous Saha equation, it is possible to express the stellar temperature in terms of the ratio of intensities of lines of HeI and HeII in the spectrum. The relationship is shown in equation (12):

$$\log R = \log\frac{N_b}{N_a} + \log\frac{N_1}{N_0}. \tag{12}$$

N_a and N_b are proportions of neutral and ionized atoms respectively in excited states producing lines.

Petrie has made use of the line λ4471 of HeI and λ4542 of HeII since these lines are situated relatively close to one another in the spectrum and their relative intensities are easy to obtain with the necessary precision. The principal difficulty is in relating the actual observed line intensities with the numbers of atoms producing the lines. This difficulty has been at least partially overcome by studies of the curve of growth for helium by Goldberg indicating the numbers of atoms producing lines of known intensity. On this basis R. M. Petrie has derived the numbers of atoms producing the lines of HeI and HeII and utilized equation (12) to determine temperatures for the O and early B type stars. His values of 28,600 for B0 and 36,300 for O5 are probably the most accurate now available for stars of this class.

TEMPERATURES OF THE EMISSION LINE STARS

Turning now to the emission line objects it may be pointed out that they offer a new approach to the problems of the temperatures of very hot stars. The first studies of stars of this type were made by Zanstra in connection with the planetary nebulae. A planetary nebula consists of a great globe of gas of as much as several light years in diameter illuminated by a star in its centre which appears always to be of very high temperature. A study of the spectra reveals the fact that in general, the spectrum of the star is continuous while that of the gaseous envelope surrounding it consists of bright lines due to atoms of high ionization potential. Among prominent lines in the spectra of planetary nebulae are those due to H, HeI, HeII, OIII and NIII.

A striking feature of the planetary nebulae as illustrated in Fig. 3, which shows the Ring Nebula in Lyra, is the large value of the integrated luminosity of the nebula as compared with that of the central star. A quantitative study of this phenomena by Zanstra showed that this could be explained as a fluorescence process in which the ultra-violet light from the hot star was absorbed by the nebula and converted into visible luminosity. The mechanism of absorption is considered to be photoelectric ionization by the atoms of the nebulae of that portion of the continuous spectrum beyond the ionizing frequency of the atom concerned. For hydrogen this is at wave-length λ912 and for atoms of higher ionization potential, it will be of correspondingly shorter wave-length. In a steady state the number of captures will equal the number of ionizations and those electrons captured in upper states will subsequently cascade to lower levels giving rise to observable emission.

The amount of radiation available for ionization will depend on the temperature of the central star and the numerical relations may be expressed in terms of the following equations due to Zanstra [26]:

$$\int_{x_0}^{\infty} \frac{x^2}{e^x - 1}\, dx = \Sigma \frac{x^3}{e^x - 1} A\nu, \tag{13}$$

where $x = \frac{h\nu}{kT}$; $x_0 = \frac{h\nu_0}{kT}$; and $A\nu$ is an observed quantity depending on the observed strength of the line. The left-hand member of this equation represents the number of quanta of ultra-violet starlight absorbed by the envelope, while the right-hand member represents the number of quanta emitted as observable emission lines.

Zanstra's theory of nebular luminosity, embodied in the above equation, represents a very important advance in our knowledge of early type stars since it makes possible estimates of the temperatures

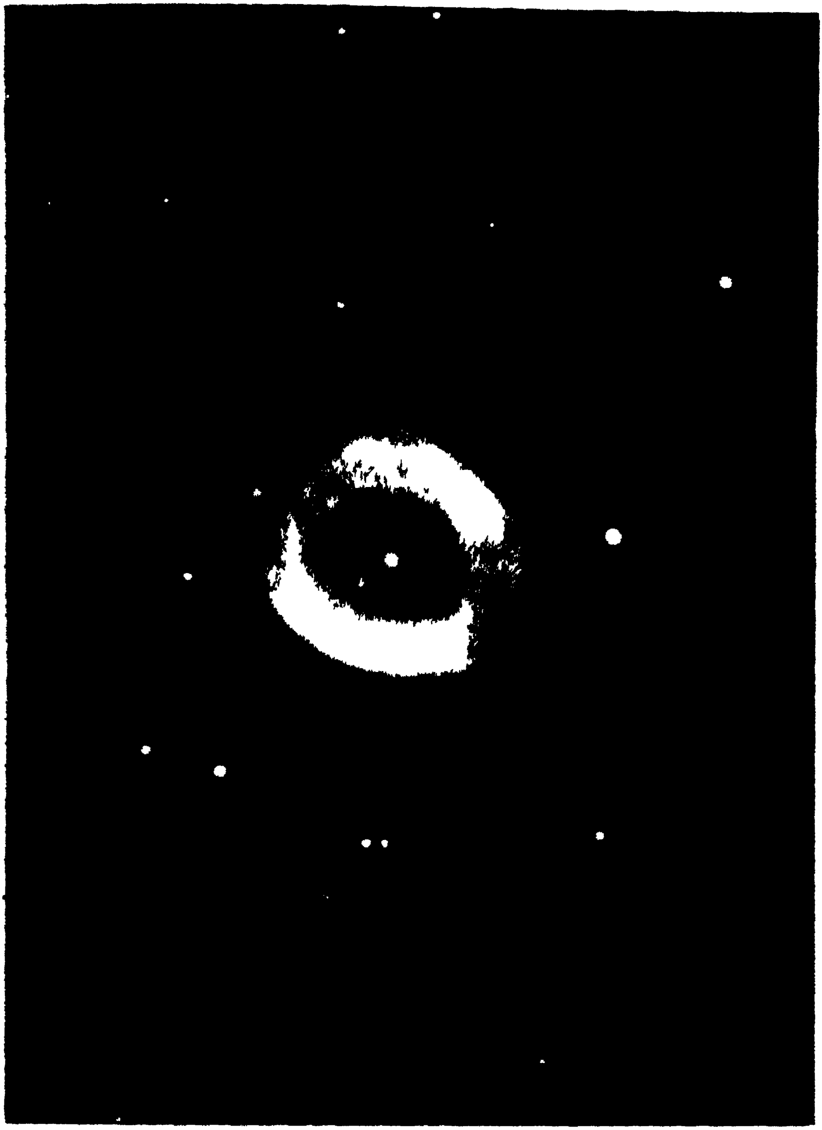


FIG. 3. The "Ring" Nebula in Lyra, photographed with the 73-inch Victoria Telescope. This ring-like object is composed of a number of shells of gas, of exceedingly low density, surrounding a very high-temperature star. The gaseous shell shown is made luminous by the absorption of energy radiated by the central star. The shell seen in this photograph is so vast that it would take light nearly a year to cross it.

of the nuclei of planetary nebulae, objects which cannot, in general, be successfully studied by other methods It will be noted that equation (13) involves the stellar temperature and that it can easily be solved by trial yielding the temperature of the central star.

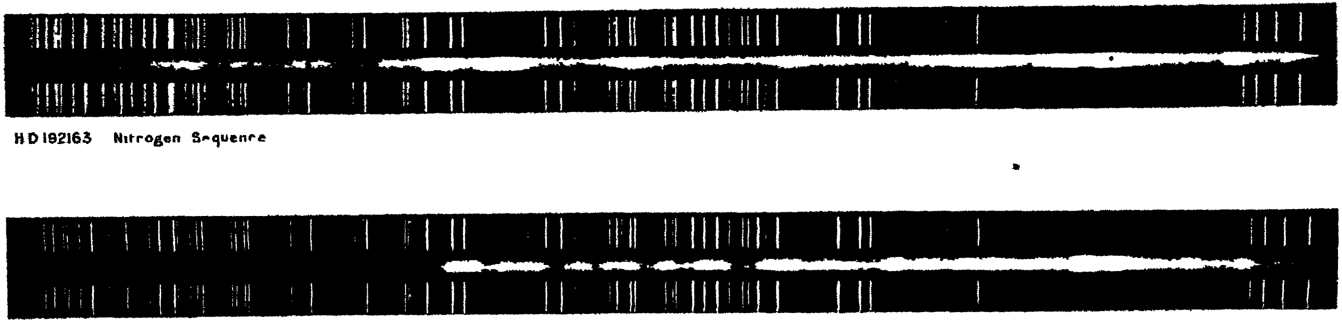
Results of the study of a number of these objects by Zanstra and others have indicated that the planetary nuclei range in temperature from 30,000° to 140,000°K.

In addition to the planetary nebulae there are many other stars which show emission lines but which are not accompanied by shells of gas large enough to be directly observed. Among stars of this type are the novae or new stars which are actually old stars which suddenly attain a new lease of life and blaze up brilliantly for a relatively brief period only to subside to their former low luminosity after a few months or years. One of the phenomena associated with a nova's rise to brightness, as has already been pointed out, is the displacement of its absorption lines to the violet by amounts corresponding to velocities of several thousand kilometres per second. Since these velocities are larger than the velocity of escape, the gases at the surface must be escaping from the star to form around it an envelope of outward moving gases which has some of the characteristics of a planetary nebula. Actually in a few cases novae have been near enough for an actual disc to be seen, giving ample confirmation to this interpretation and it might reasonably be expected that if the temperature of the central star is sufficiently high, emission lines would be observed.

Actually emission lines are a conspicuous feature of all nova spectra and they increase in intensity after maximum light when the nova is known to increase in temperature. There now appears to be no question but that the origin of these lines is similar to those of planetary nebulae and the relation between line intensity and the temperature of the exciting star can be derived by Zanstra's equation quoted above. The use of this equation makes it possible to follow the temperature of the nova from around 20,000° near maximum light to 100,000° when the emission bands reach their maximum intensity.

In addition to the novae there are numerous other stars which exhibit emission lines or bands but which do not exhibit major changes in luminosity of the kind shown by the novae. The Wolf Rayet stars, spectra of which are shown in Fig. 4, are characterized by broad emission lines of great intensity due to highly ionized atoms and the emission bands are accompanied by absorption borders on their violet edges. It appears, therefore, that these stars also are surrounded by envelopes of ejected gases and that their temperatures may be derived by the methods already mentioned. Such determinations indicate

EMISSION BANDS IN TWO WOLF RAYET STARS



HD 192163 Nitrogen Sequence

HD 192103 Carbon Sequence

FIG 4 Emission bands in two Wolf Rayet stars The Wolf Rayet stars (discovered by Messrs Wolf and Rayet at the Paris Observatory in 1867) are remarkable for the bright emission bands which appear in their spectra Studies of these stars made at Victoria have indicated that the bright bands are due to an envelope of gases surrounding the star This envelope is caused by the ejection of gases from the star by radiation pressure with velocities sometimes as great as 4,000 kilometres per second These envelopes are not uniform in chemical composition, some showing an excess of carbon and oxygen while others show mainly nitrogen

temperatures for the Wolf Rayet stars ranging from 50,000°K. to 100,000°K.

The P Cygni stars, profiles of which are illustrated in Fig. 5, are in many ways similar to those of the Wolf Rayet type except that the displaced absorption components are more intense relative to the emission, and the velocities of ejection are much less, of the order of a few hundred kilometres per second. The application of Zanstra's equation to the emission line intensities leads to values of temperature ranging from 10,000° to 42,000°.

In addition to the types of stars mentioned above whose spectra give definite evidence of the type of motion required to produce an

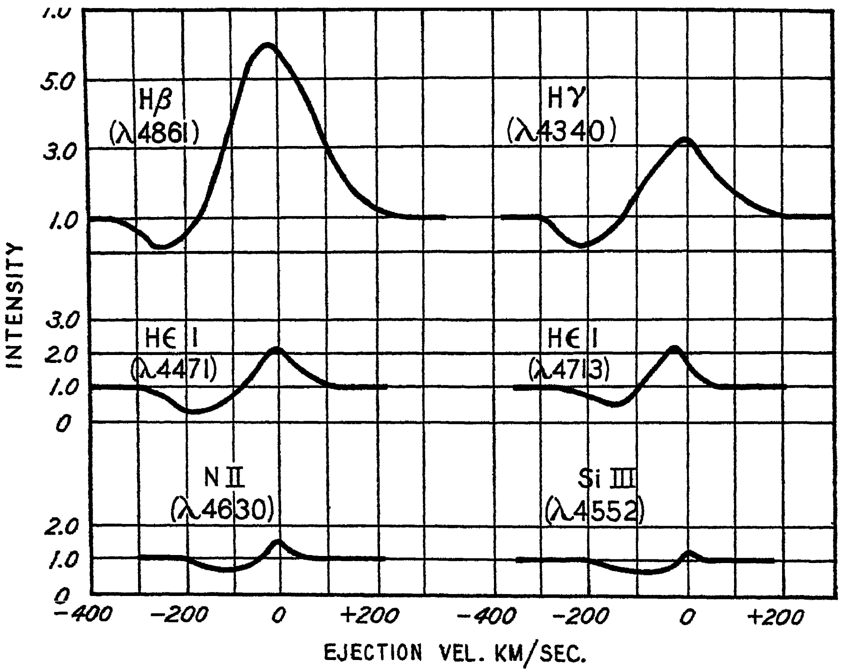


FIG. 5. Intensity profiles of P Cygni lines. In this figure intensity is plotted against velocity for a number of lines in the spectrum of the star P Cygni. The figure illustrates the relative intensities of absorption and emission components and the difference in the ejection velocity from line to line.

envelope of gases around them, there are numerous other early type stars which also show emission lines, and spectra of two stars of this type are illustrated in Fig. 6. Mainly by analogy with the emission objects already described, it is now generally believed that the emission lines in these objects also arise in an envelope of diffuse gases surrounding the star. Fortunately in one or two instances of eclipsing stars there is independent evidence supporting the view. For the star R. W. Tauri, Joy [27] has observed that the spectrum is ordinarily of absorption type, but at times of total eclipse of the brighter component by a cooler companion, emission lines appear conspicuously indicating the presence of a gaseous envelope of considerably greater diameter than the primary star. In addition, certain of these objects during the passage into or out of total eclipse, exhibit rotation effects associated with the emission lines which are consistent with the idea that the emission lines originate in a diffuse envelope which rotates with the star.

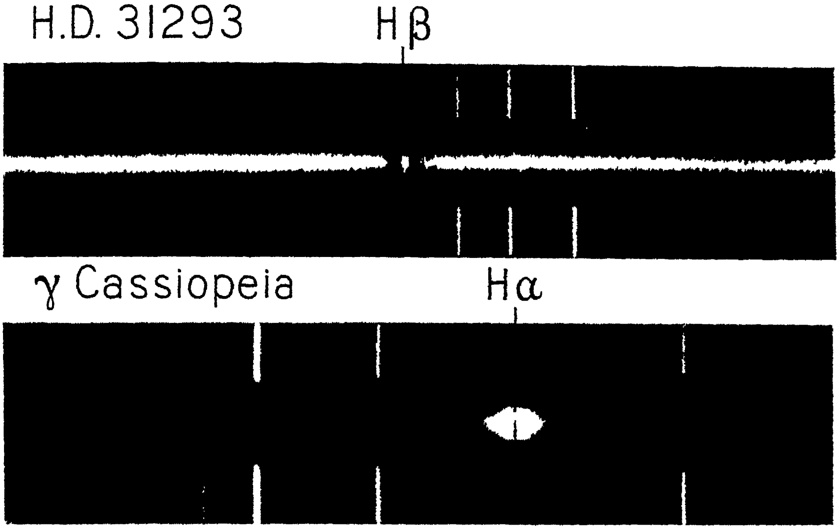


FIG. 6. Emission lines in B$_e$ stars. These spectra show two of the characteristic features of emission lines of this type, e.g., the appearance of an emission line superimposed on broad absorption (H.D. 31293) and a relatively broad emission line cut by sharp absorption (γCaseopeia).

For many years, the origin of emission lines in early type stars was an unsolved mystery and the recognition that such lines are due to an extended gaseous envelope illuminated by a hot star is in itself an interesting development in astrophysical science. Of considerably greater importance, however, is the possibility of shedding additional light on the nature of absorption line stars. The main effort in astrophysical research has always been concentrated on these more normal objects since they outnumber the emission line stars by approximately one thousand to one. The emission line stars have sometimes been referred to as pathological cases with the implication that astrophysical results derived from them are of less general validity or application than for the more normal absorption line objects.

While this may be so, it may be pointed out that the study of abnormal psychology has often led to advances in knowledge concerning the mental processes of normal human beings. Similarly with stars, the study of the abnormal emission line objects may lead to valuable information concerning the stellar sequence as a whole, and I should like to conclude this discussion by drawing your attention to an instance of this character.

A good many otherwise normal absorption line stars have a few emission lines associated with them while others, especially among the P Cygni stars with conspicuous emission, have sufficiently well-developed absorption line spectra to place them with reasonable definiteness in their proper place in the Harvard system of classification. For these stars as well as for others of more completely developed emission line characteristics, it is possible to utilize measured emission line intensities to determine stellar temperatures on the basis of Zanstra's equation. Since sporadic emission lines are found in all the early type classes down to and including the early A's, it is possible to draw up a scale of temperatures for the entire range of early type stars based on emission line character alone.

Such a temperature scale is shown in Table I. Here Zanstra's temperatures are used for the planetary nebulae, the results of Beals are used for the Wolf Rayet stars, while values derived by Beals and Hatcher [28] are used for the P Cygni stars which cover the O, B and early A stars. For purposes of comparison, the table also includes the ionization temperatures derived by R. M. Petrie for the absorption O's and the colour temperatures of the B's and early A's by the Greenwich observers [29]. The agreement is generally very good and suggests that we now have for the early type stars as a whole a temperature scale in which we may have reasonable confidence.

One of the aspects of this table which cannot fail to attract attention is the position occupied by W stars and the planetary nebulae relative to the other early type stars. Considered on the basis of temperature alone it would appear that these interesting emission line objects simply represent an extension towards earlier spectral types of the normal Harvard sequence of absorption line stars. The possibility that this may actually be the case receives support from the fact that no absorption line stars have been definitely observed in the temperature range from 40,000° to 100,000°K. It is also well known that the tendency for the stars to exhibit emission line spectra is greater for

TABLE I

TEMPERATURE SCALE
EARLY TYPE STARS

Class	*Temperature*		
	Absorption Lines and Continuous Spectrum	Emission Lines	Adopted
Planetary nebulae		140,000 to 30,000	140,000 to 30,000
W 5 W 8		110,000 50,000	110,000 50,000
O 5 O 9	36,300 30,700	40,000 34,000	38,000 32,000
B 0 B 9	25,000 13,500	32,500 17,000	28,000 15,000
A 0 A 4	14,000 11,000	16,000 11,000	14,000 11,000

increasing temperature. Although the absolute number of emission line objects is greater for the B's, the proportionate number is greater in the O class. If this proportion of emission to absorption line objects changes with sufficient abruptness at about O5 in the sequence, then it may be permissible to regard the W stars as a normal part of the Harvard sequence and their emission line character as a consequence largely, if not entirely, of their higher temperatures.

For temperatures above 50,000°, therefore, it is suggested that the normal type of stellar spectrum may be of emission type just as for lower temperatures the normal spectrum is of absorption line character.

As for the planetary nuclei it may be pointed out that some of them appear to be ordinary O type stars, while others exhibit Wolf Rayet bands in no way different from other W stars. There are, however, a number of the hotter planetary nuclei which, in so far as the writer is aware, have neither Wolf Rayet bands or absorption lines in their spectra. The strong emission lines which characterize the spectra of the gaseous nebulae which surround them make it difficult to be certain of the presence or absence of absorption lines, but present indications are that they are either not present or are of very low intensity.

Little that is definite can be said at present about the interesting question as to their relation to the Wolf Rayet stars. It appears, however, that the two types of objects are roughly parallel as regards temperature and it may be that the relations between them are in a general way similar to those existing between the absorption and emission line objects in the O and B classes. The writer believes that the most helpful approach to the problem at the present time is tentatively to regard the entire group as a normal extension of the Harvard sequence and to see whether the methods of theoretical astrophysics may not give an indication of the types of spectra to be expected for a reasonable range of masses, diameters, and chemical compositions. Some interesting suggestions along this line have recently been given by Miss Anne Underhill [30], who has indicated the probability that the effects of radiation pressure will render unstable the atmospheres of most stars of temperature greater than 50,000°. It seems possible that a further application of the same methods might shed some light on the origin of the observed differences between the spectra of the W stars and the nuclei of the hotter planetaries.

REFERENCES

[1] J. Homer Lane, Amer. J. Sci. 2nd series, vol. 50 (1869), 57-74.
[2] R. Emden, *Gaskugeln* (Leipzig, 1907).
[3] A. S. Eddington, *The internal constitution of the stars* (Cambridge University Press, 1926).
[4] J. H. Jeans, *Problems of cosmogony and stellar dynamics* (Cambridge University Press, 1919).
[5] S. Chandrasekhar, *An introduction to the study of stellar structure* (University of Chicago Press, 1939).
[6] R. M. Petrie, Pub. Dom. Ap. O., vol. 8, in press.
[7] A. Unsold, *Physik der Sternatmospharen*, p. 286 (Springer, Berlin, 1938).
[8] B. Holtzmark, Ann. d. Physik, vol. 58 (1919), 576.
[9] Otto Struve, Observatory, vol. 61 (1938), 53.
[10] J. S. Foster and A. V. Douglas, Monthly Notices, R.A.S., vol. 99 (1939), 150.
[11] Anne Underhill and W. Petrie, J. Roy. Astr. Soc. Can., vol. 38 (1944), 385.
[12] Anne Underhill, Ap. J., vol. 107 (1948), 349.
[13] Otto Struve, Ap. J., vol. 72 (1930), 1.
[14] J. Carroll, Monthly Notices, R.A.S., vol. 93 (1933), 478.
[15] Otto Struve, Ap. J., vol. 104 (1946), 138.
[16] H. D. Babcock, Ap. J., vol. 105 (1947), 105, and Pub. Astr. Soc. Pac., vol. 59 (1947), 260.

[17] L. Campbell and L. Jacchia, *The story of variable stars* (Harvard Books of Astronomy, 1941).

[18] W. E. Harper, J. A. Pearce, C. S. Beals, R. M. Petrie, and A. McKellar, Pub. Dom. Ap. O., vol. 6 (1937), 317

[19] C. S. Beals, Pub. Dom. Ap. O., vol. 4 (1930), 271, and vol. 6 (1934), 93

[20] C. S. Beals, J. R.A.S.C., vol. 34 (1940), 169; vol. 36 (1940), 145; and Observatory, vol. 64 (1941), 42; vol. 57 (1934), 319

[21] W. Petrie, Pub. Dom. Ap. O., vol. 7 (1947), 383.

[22] M. Saha, Phil. Mag., vol. 40 (1920), 472, 809.

[23] R. H. Fowler and E. A. Milne, Monthly Notices, R.A.S, vol. 83 (1923), 403; vol. 84 (1924), 499.

[24] Cecilia Payne, *Stellar atmospheres* (Harv. Mon, 1925).

[25] R. M. Petrie, Pub. Dom. Ap. O., vol. 7 (1947), 334.

[26] H. Zanstra, Pub. Dom. Ap. O., vol. 4 (1931), 209

[27] A. H. Joy, Pub. Astr. Soc. Pac., vol. 59 (1947), 171.

[28] C. S. Beals and R. D. Hatcher, Contr. Dom. Obs., vol. 2, No. 1, 1948.

[29] F. Dyson, *Colour temperatures of stars* (Greenwich, H.M. Stationery Office, 1932).

[30] Anne Underhill, Monthly Notices, R.A.S., vol. 109 (1949), 562.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLIV : SERIES III : JUNE, 1950
SECTION THREE

The Action of Hydrogen Peroxide—Nitric Acid Mixtures on Ketones*

By G. H BJORKLUND and W H. HATCHER, F R S C

THE observations here recorded are a direct result of researches conducted on the action of hydrogen peroxide with and without the presence of mineral acids Holden [5] found that traces of these acids (which in quantity produced considerable side-reactions) were mainly catalytic. Sturrock [15] and Mueller [11] demonstrated that hydrogen peroxide and various organic acids formed first a complex, which might give rise to a peracid or decompose into simpler products.

The researches of Maass and coworkers [6], [7], [8], [9] and of MacLauchlan [10] correlated the effects of conductivity and concentration and set limits beyond which the latter produced serious decomposition of the peroxide. In this connection the work of Caro [2] and Raschig [13] takes on added significance. Eventually it became necessary to study the oxidative reaction of nitric acid–hydrogen peroxide mixtures on some well known carbon compound; that chosen was acetone, whose reaction with hydrogen peroxide alone had been investigated by Wolffenstein [17] and by Baeyer and Villiger [1]. The results obtained suggested extension of this reaction to other ketones.

Reagents.—The acetone used was of research grade and showed no decolourization with neutral permanganate solution. The nitric acid was of C.P. grade, nitrite-free and of 70.0 per cent by weight. The hydrogen peroxide was specially prepared without inhibitor and originally 30 per cent by weight.

Experimental.—Considerable time was spent in discovering in what concentrations the reagents could be mixed so that the subsequent reaction could be controlled. For example, a solution showing the following molar concentrations—acetone 0.23, peroxide 0.23, nitric acid 0.65, water 2.10—got out of hand at room temperature and exploded with great violence; whereas the same concentrations, carefully controlled as to temperature, gave a 98 per cent yield calculated on the amount of acetone used.

The procedure finally adopted was as follows. Aqueous solutions of hydrogen peroxide (27.5 per cent) and nitric acid (70.0 per cent) were cooled to 5°C. and then mixed in desired proportions. This mixture was kept at 50°C. until all heating effects had subsided; to this was added the appropriate volume of acetone, also at 5°C., and the bath kept at 5°C. Under these circumstances a satisfactory yield of solid was obtained immediately and the reaction temperature could be kept as low as 33°C. If the temperature exceeded 50°C. at this time the reaction was very liable to progress rapidly to 90°C. with violent evolution of heat and loss of reagents. Small beakers were found to be the most suitable containers. The molar concentrations studied ranged as follows: acetone, 0.23 to 0.52; hydrogen peroxide, 0.23 to 0.43; nitric acid, 0.25 to 0.65; water, 2.1 to 2.5. A mixture containing more than 40 per cent by weight of nitric acid is potentially dangerous. The above concentrations were those which offered the least hazard.

---

* Presented as partial fulfilment of the requirements for the degree of Ph.D., McGill University.

In general it may be said that the yield of product increased as the concentration of nitric acid present. Also, a higher yield was obtained by keeping the temperature low at the same time.

Two different solids were obtained from these mixtures. Under a given set of conditions, only one product appeared rather than a mixture of the two. Where the acid and hydrogen peroxide were added first, the subsequent addition of acetone favoured the formation of the product of high melting point. When the acetone and peroxide were mixed first and the acid added later, the other product tended to form first. Further, a low temperature under either mode of addition favoured the formation of the low melting product, whereas an excess of acid favoured the high melting product. In any case, precipitation was immediate.

In both cases the product when first formed is white and apparently amorphous. One melted at 95°-97°C. and corresponds in every way to the tricyclo-acetone peroxide of Wolffenstein.

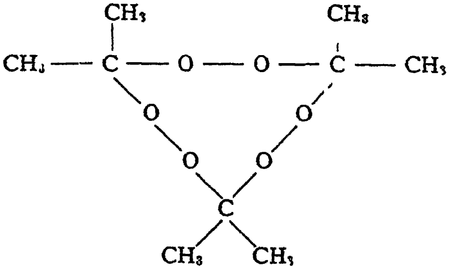
The other product had a melting point of 131°-132°C. and resembles the dicycloacetone peroxide of Baeyer and Villiger.

Both compounds are insoluble in water, dilute acids, and alkalies; but hydrolysis in dilute sulphuric acid proceeds slowly to regenerate acetone and hydrogen peroxide. They explode violently under impact, when quickly heated and when mixed with concentrated sulphuric acid. In this respect the lower melting compound is the more violent. Both evaporate slowly when exposed to the air. They are insoluble in alcohol but soluble in ether, ethyl acetate, and petroleum ether.

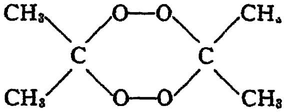
Because of this sensitivity, it was not deemed advisable to perform a combustion analysis. But cryoscopic results from benzene showed the following molecular weights:

I. $(C_3H_6O_2)_3$ of m. pt. 95°-7°C.; found 216, 227;
Calc. 222.
II. $(C_3H_6O_2)_2$ of m. pt. 131°-2°C.; found 142:
Calc. 148.

To I, discovered by Wolffenstein, Baeyer and Villiger had assigned the formula



They suggested for II:



which they had discovered.

The close similarity in chemical properties suggests a similar structure, of which more will be said later.

Other ketones—Methyl ethyl and di-ethyl ketones. Baeyer and Villiger [1] had treated the former with Caro's acid and found an explosive peroxide.

Solutions of each of the above were made up to contain 0.1 mole of ketone, 0.14 mole of hydrogen peroxide, and 0.44 mole of nitric acid, the usual precautions being taken. After a slight rise in temperature an oily layer separated out on top of each liquid. These layers were taken up in ether, neutralized with sodium bicarbonate, dried over anhydrous sodium sulphate and recovered by evaporation of the solvent.

Both oils were colourless and of not unpleasant odours; they were highly sensitive to shock and sudden heating. Their solubilities are similar to those of the acetone peroxides.

Di-isopropyl ketone.—A similar treatment of this ketone showed no reaction whatever over a period of three months.

Cyclopentanone.—This ketone reacts very vigorously with the acid mixture. Ketone, 0.3 mole, was added to an acid mixture made up of 0.58 mole of hydrogen peroxide and 0.7 mole of nitric acid and the mixture cooled well below 10°C. A large quantity of white amorphous material was precipitated, which, after about five minutes, produced a series of small explosions. This was arrested by adding a large volume of cold water, and immediately filtering and washing the solid.

Crystallization was effected from ethyl acetate, and the compound decomposed at 166°-167°C. This peroxide is very sensitive to shock, slight friction, and quick heating. It explodes very violently.

A molecular weight determination was made as previously described:

$(C_5H_8O_2)_3$ Found 275 Calc. 284.

Cyclohexanone.—Cyclohexanone produced rather inconsistent results. In some experiments a resinous product was accompanied by a small amount of a solid, both explosive. Other experiments gave rise to a viscous oil, also explosive, but to no precipitate. This oil was removed and treated with methyl alcohol, when a solid was obtained. This crystallized from ethyl acetate, and gave a melting point of 126°-127°C. Its explosive properties are similar to those in the previous case, although, in general, not quite so vigorous.

A molecular weight determination showed

$(C_6H_{10}O_2)_2$ Found 232 Calc. 228.

m-Methyl cyclohexanone.—This ketone, when treated with a hydrogen peroxide–nitric acid mixture, gave rise to an oil which could not be crystallized. It, too, proved to be explosive.

Acetoacetic ester.—This was easily soluble in the acid-peroxide mixture and after 25 days gave rise to a product identified with dicycloacetone peroxide.

Urea and thiourea.—Though the former showed no noticeable reaction, the latter, when dissolved first in glacial acetic acid, produced a considerable precipitate. This, on drying in the air, decomposed like miniature volcanoes, evolving sulphur dioxide and scattering free sulphur about.

Benzil.—Pastureau [12] claims that hydrogen peroxide in sulphuric acid cleaves benzil. Weitz and Schaffer [16] observed that alkaline hydrogen peroxide cleaves 1, 2-diketones very readily.

Benzil reacted very slowly with the acid mixture. After a period of seven days a small amount of solid separated out; this was identified as benzoic acid.

Beta naphthol.—This reacted slowly to produce a series of coloured solutions, the only identifiable product being probably o-carboxy cinnamic acid or a rearrangement of it (Roth [14]). Its amide melted at 198°-199°C. and its silver salt was dark red.

Hydrolysis of ketone peroxides.—Many tests were made by allowing these substances to stand in dilute sulphuric acid. In each case hydrogen peroxide was regenerated but at a very slow rate, probably because of their relative insolubilities.

Addition of peroxides to aqueous potassium iodide produces iodine very slowly. Some tests were followed over a period of weeks by means of $N/50Na_2S_2O_3$. Since soluble peroxides usually liberate iodine much more slowly than peracids, such rates as shown here may be caused by insolubility and rendered unreliable by virtue of the simultaneous evolution of hydrogen peroxide.

Conclusions.—The normal reaction of hydrogen peroxide with water-soluble ketones results in the formation of compounds where two or more molecules of ketone unite with oxygen atoms in the ratio of 1: 1. The presence of controllable amounts of nitric acid greatly accelerates such addition, provided that the temperature of the mixture is held below 40°C.

The formation of such compounds is undoubtedly caused by some constituent which may be an isomer of nitric acid (Gleu and Hubold [4]) or a so-called pernitric acid HNO_4 (Raschig [13]) or nitrogen heptoxide (Friend [3]).

The results presented above are in accord with the reaction of hydrogen peroxide on acetaldehyde (Baeyer and Villiger [1]) and indicate that hydrogen peroxide in strongly acid solution is not a useful oxidizing agent when compared with its action in the presence of alkali where oxygen is liberated catalytically.

Indeed it is possible to raise the question of the true identity of peracids. Most of these are obtained by the direct addition of hydrogen peroxide to the acid: their production may be immediate as in the case of Caro's acid or slow as in the case of the organic acids. Perhaps they should be regarded as peroxides of the acids where the oxygen is more loosely held than in the peroxides, as determined by the evolution of iodine from potassium iodide.

The peroxides, though isolable, are unstable to shock and sudden heating; the peracids are not isolable, apparently, nor are their salts capable of a separate existence. They do not resemble the stable, strong peracids such as $HClO_4$ or $H_2S_2O_8$ which are not oxidizing agents. Whatever the exact structure of peracids and peroxides, the prefix "per" is misleading.

The formulae proposed by Baeyer and Villiger for the acetone peroxides are unsatisfactory, and such formulae are equally inapplicable to the other peroxides here studied.

REFERENCES

[1] V. Baeyer and V. Villiger. Ber., vol. 32 (3) (1899), 3625, vol. 33 (1) (1900) 126.

[2] N. Caro. Zeit. angew. Chem., vol. 11 (1898), 845.

[3] J. N. Friend. *Textbook of inorganic chemistry*, vol. 6 (1928), 202.

[4] K. Gleu and R. Hubold. Zeit. anorg. allgemein. Chem. (1935), 223, 305.

[5] G. W. Holden and W. H. Hatcher. Trans. Roy. Soc. Can., vol. 17 (3) (1923), 119; vol. 18 (3) (1924), 231; vol. 21 (3) (1937), 237.

[6] O. Maass and P. Hiebert. J. Amer Chem. Soc., vol. 46 (1920), 290.

[7] ——— and A. C. Cuthbertson. J. Amer. Chem. Soc., vol. 52 (1930), 489.

[8] ——— and E. P. Linton. J. Amer. Chem. Soc., vol. 53 (1931), 957.

[9] ——— and G. L. Matheson. J. Amer. Chem. Soc., vol. 51 (1929), 674.

[10] D W. MacLauchlan and W. H. Hatcher. Can. J. Res., vol 16B (1938), 253.

[11] W. H. Mueller and W. H. Hatcher. Can. J. Res, vol. 3 (1930), 291.

[12] M. Pastureau. Bull soc. chim, vol. 5 (4) (1909), 227.

[13] F. Raschig. Zeit. angew. Chem, vol. 17 (1904), 1419.

[14] R. Roth. Ber., vol. 47 (2) (1914), 1597.

[15] M. G. Sturrock and W. H. Hatcher. Can. J. Res, vol. 4 (1931), 35.

[16] Weitz and Schaffer. Ber., vol. 54 (1921), 2327.

[17] R. Wolffenstein. Ber., vol. 28 (1895), 1459.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLIV : SERIES III : JUNE, 1950
SECTION THREE

The Steinitz-Gross Theorem on Sums of Vectors

By IRA DAMSTEEG and ISRAEL HALPERIN

Presented by R. L. JEFFERY, F.R.S.C.

1. Introduction. If a series of real numbers is convergent but not absolutely convergent, then under different ways of ordering its terms the series can converge to different sums. Indeed, the set of such sums, as Riemann showed, consists of *all* real numbers. In a generalization to series of vectors, Steinitz [1] proved:

THEOREM A. *If the series of m-dimensional vectors* $\sum_{n=1}^{\infty} v_n$ *is convergent but not absolutely convergent, then its sums (allowing all orderings of the terms) form a linear subspace.*

Thus, to take the 2-dimensional case, if a series of complex numbers is convergent but not absolutely convergent, then under rearrangement of the terms, either the sums are precisely all the points of some straight line in the complex plane or they include every complex number.

Steinitz based part of his proof of Theorem A on essentially the following theorem:

THEOREM B. *If* $v_n \to 0$ *as* $n \to \infty$ *and the infinite series* $(v_1 + v_2 + \dots + v_{n_1}) + (v_{n_1+1} + \dots + v_{n_2}) + \dots$ *converges, then within each bracket the terms can be rearranged so that the series will converge with the brackets removed.*

Theorem B was deduced easily from a theorem which can be stated as follows:

THEOREM C. *For* $c_1 \geqslant 0$, $c_2 \geqslant 0$ *and positive integer m there is a finite constant* $K_m(c_1, c_2)$ *with the property: whenever, for any n, the m-dimensional vectors* $u_1, v_1, \dots, v_n, u_2$ *satisfy* $|u_1| \leqslant c_1$, $|u_2| \leqslant c_2$, $|v_i| \leqslant 1$ *for all i, and* $u_1 + v_1 + \dots + v_n + u_2 = 0$ *then by reordering the* v_i *it is possible to satisfy*

$$\left| u_1 + \sum_{i=1}^{h} v_i \right| \leqslant K_m(c_1, c_2) \text{ for } h = 1, 2, \dots, n.$$

In what follows, $K_m(c_1, c_2)$ shall denote the least possible such constant. For given m, its finiteness for any particular c_1, c_2 obviously implies its finiteness for all c_1, c_2. Also, $K_m(c_1, c_2) = K_m(c_2, c_1)$ as can be seen by reversing the order in each arrangement of the given vectors.

By a detailed analysis Steinitz showed that $K_m(0, 0) \leqslant 2m$ from which it follows immediately that $K_m(1, 1) \leqslant 2(m + 1)$. Later Gross [2] gave a simple proof by induction which showed that $K_m(1, 1) \leqslant 2^m - 1$.

Gross posed the problem of determining $K_m(0, 0)$, $K_m(1, 0)$ and $K_m(1, 1)$ precisely and proved that $K_2(1, 0) = K_2(1, 1) = \sqrt{2}$. First he showed by an example that $K_2(1, 0) \geqslant \sqrt{2}$ (and hence that $K_2(1, 1) \geqslant \sqrt{2}$). Then he gave two proofs, one that $K_2(1, 1) \leqslant \sqrt{2}$ and the other that $K_2(1, 0) \leqslant \sqrt{2}$. Both proofs are involved but particularly so is the first one which gives the stronger result. In section 2 we give a simple proof that $K_2(1, 1) \leqslant \sqrt{2}$ which we obtained when we knew Gross's result but not his methods. It turns out that our method is related to his proof that $K_2(1, 0) \leqslant \sqrt{2}$.

In section 3 we give examples to show that for all m, $K_m(0, 0) \geqslant \frac{1}{2}(m + 3)^{\frac{1}{2}}$, and that for all $m \geqslant 2$, $K_m(1, 0) \geqslant \frac{1}{2}(m + 6)^{\frac{1}{2}}$ (and *a fortiori*, $K_m(1, 1) \geqslant \frac{1}{2}(m + 6)^{\frac{1}{2}}$). These inequalities show that Theorem B is not valid for series of vectors in Hilbert space so that the methods of Steinitz and Gross do not serve to extend Theorem A to such series.

2. In this section all points and vectors will lie in a fixed plane, $|w|$ will denote the length of the vector w, all v_i will be vectors with $|v_i| \leqslant 1$ and 0 will be the centre of a fixed circle C of unit radius. PQ will denote the vector from P to Q.

We introduce the concepts of "special position" and "preferred vector" as follows:

Definition 1. A point P_0 is in special position relative to vectors $v_1, \ldots, v_n$ if

(i) $1 < |OP_0| \leqslant \sqrt{2}$

(ii) $|OP_0 + v_1 + \ldots + v_n| \leqslant 1$

(iii) there is a tangent TP_0T' to the circle C, with T on the circle, such that none of the v_i point from P_0 into the interior of the angle bounded by the vectors P_0O and P_0T'.

Definition 2. Let θ_i designate the angle between the vector v_i and P_0O and $|\theta_i|$ its numerical, non-negative value with the convention that the zero vector makes zero angle with every vector. Then a particular v_j is called preferred relative to $P_0, v_1, \ldots, v_n$ if $|\theta_j| \leqslant |\theta_i|$ for $i = 1, \ldots, n$.

Under the assumptions that $v_1 = P_0P_1$ is preferred relative to P_0, $v_1, \ldots, v_n$ and that $|OP_0 + v_1 + \ldots + v_n| \leqslant 1$ we shall prove below the following two lemmas:

LEMMA 1. If P_0 is in or on the circle C then P_1 is either in or on the circle C or in special position relative to $v_2, \ldots, v_n$.

LEMMA 2. If P_0 is in special position relative to $v_1, \ldots, v_n$ then P_1 is either in or on the circle C or in special position relative to $v_2, \ldots, v_n$.

Suppose now that $u_1 = OP_0$ and $v_1, \ldots v_n$ are given vectors with $u_1| \leqslant 1$ and $|u_1 + v_1 + \ldots + v_n| \leqslant 1$ and let P_i be defined by $OP_i = u_1 + v_1 + \ldots + v_i$. Re-order the v_i so that each v_i is preferred relative to $P_{i-1}, v_i, v_{i+1}, \ldots, v_n$. Then Lemmas 1 and 2 show, by induction on i, that each P_i is either in or on C or in special position relative to some set of vectors. Hence $|OP_i| \leqslant \sqrt{2}$ for all i, showing that $K_2(1, 1) \leqslant \sqrt{2}$.

Proof of Lemma 1. We may suppose that $|OP_1| > 1$ and need only prove that P_1 is in special position relative to $v_2, \ldots, v_n$. Since $|P_0P_1| \leqslant 1$ it follows that $P_0 \neq 0$ and we may choose rectangular x, y axes with origin at 0, with positive x axis along OP_0, so that P_1 has non-negative y-coordinate. Let P_0 be $(x_0, 0)$ and P_1 be (x_1, y_1) so that $0 < x_0 \leqslant 1$, $0 \leqslant y_1$. The inequalities $|OP_1| > 1$, $|P_0P_1| \leqslant 1$ imply that $x_1 > 0$.

Now $|x_1| = x_1 \leqslant 1$ for if $x_1 \leqslant x_0$ then $x_1 \leqslant x_0 \leqslant 1$; if $x_1 > x_0$ then each v_i has positive x-component since v_1 is preferred, and hence $x_1 = x$-component of $OP_0 + v_1 \leqslant x$-component of $OP_0 + v_1 + \ldots + v_n \leqslant 1$. Also $|y_1| \leqslant |P_0P_1| \leqslant 1$. Hence $|OP_1| \leqslant \sqrt{2}$ so that condition (i) of Definition 1 is satisfied. Incidentally, the inequalities $|OP_1| > 1$ and $|x_1| \leqslant 1$ imply that $y_1 > 0$.

Condition (ii) of Definition 1 is satisfied since $|OP_1 + v_2 + \ldots + v_n| = |OP_0 + v_1 + \ldots + v_n| \leqslant 1$.

Since $0 < x_1 \leqslant 1$, $0 < y_1 \leqslant 1$ there is a unique tangent line $T_1P_1T'_1$ with point of contact $T_1 = (\bar{x}_1, \bar{y}_1)$ such that $x_1 \leqslant \bar{x}_1 \leqslant 1$ and $0 \leqslant \bar{y}_1 \leqslant y_1$. Let P_0O_0 be drawn parallel to P_1O. Since $|OP_1| > |P_0P_1|$, the angle bounded by P_0O and P_0P_1 is greater than that bounded by P_0O and P_0O_0. Hence none of $v_2, \ldots, v_n$ point into the interior of the angle bounded by P_0O_0 and P_0P_1 since v_1 was preferred relative to $P_0, v_1, \ldots, v_n$. *A fortiori*, none of $v_2, \ldots, v_n$ point into the interior of the (included) angle bounded by $P_1T'_1$ and P_1O. This proves that condition (iii) of Definition 1 is satisfied.

Proof of Lemma 2. We may choose rectangular x, y axes with origin at O so that P_0 has coordinates $(1, y_0)$ with $0 < y_0 \leqslant 1$ and so that none of the v_i point from P_0 into the angle bounded by P_0O and $P_0T'_0$ where $P_0T'_0$ is a vector with components $[0,1]$. Let P_1 be (x_1, y_1). Since v_1 is preferred, $x_1 > 1$ is impossible, for this would imply that the

x-component of every v_i is positive and hence that the x-component of $OP_0 + v_1 + \dots + v_n$ is greater than 1, which is false. Thus P_1 must lie in the sector bounded by the ray $P_0 O$ and the ray opposite to $P_0 T'_0$. Now from $|P_0P_1| \leqslant 1$ it follows that $0 < x_1 \leqslant 1$, and $-1 \leqslant y_1 \leqslant y_0 \leqslant 1$. We may suppose that $|OP_1| > 1$ and need only show that P_1 is in special position relative to $v_2, \dots, v_n$.

Since $|x_1| \leqslant 1$ and $|y_1| \leqslant 1$ it follows that condition (i) of Definition 1 is satisfied.

Since $|OP_1 + v_2 + \dots + v_n| = |OP_0 + v_1 + \dots + v_n| \leqslant 1$ it follows that condition (ii) of Definition 1 is satisfied.

Since $|OP_1| > 1$ but $|x_1| \leqslant 1$ it follows that $|y_1| \neq 0$. We may consider the two cases $y_1 > 0$, $y_1 < 0$ separately. In either case there is a unique tangent line $T_1P_1T'_1$ with point of contact $T_1 = (\bar{x}_1, \bar{y}_1)$ such that $x_1 \leqslant \bar{x}_1 \leqslant 1$ and $|\bar{y}_1| \leqslant |y_1|$. The angle bounded by the vectors $P_1T'_1$ and P_1O is contained in the angle bounded by P_0P_1 and $P_0T'_0$. Since none of $v_2, \dots, v_n$ point into the interior of the latter angle, condition (iii) of Definition 1 is verified.

3. The examples in m-dimensional space discussed in this section give lower bounds for $K_m(0, 0)$, $K_m(1, 0)$ and $K_m(1, 1)$ which show that these constants become infinite as m becomes infinite.

Let $e_1, \dots, e_m$ be mutually orthogonal unit vectors, $N > \frac{\sqrt{m-1}}{4}$, $a = \sqrt{1 - \frac{m-1}{16N^2}}$ and define $v_1, \dots, v_{m+4N-1}$ as follows:

$$v_i = e_i \qquad \text{for } 1 \leqslant i < m.$$

$$v_i = ae_m - \frac{e_1 + \dots + e_{m-1}}{4N} \qquad \text{for } m \leqslant i < m + 2N.$$

$$v_i = -ae_m - \frac{e_1 + \dots + e_{m-1}}{4N} \qquad \text{for } m + 2N \leqslant i < m + 4N.$$

Then $|v_i| = 1$ for all i and $v_1 + \dots + v_{m+4N-1} = 0$. Yet for every arrangement of the v_i, either the first partial sum which contains $2N$ of the vectors $v_m, \dots, v_{m+4N-1}$ or the first partial sum which contains $2N + 1$ of these vectors has length not less than

$$\sqrt{a^2 + (m-1)\frac{(2N-1)^2}{16N^2}} = \sqrt{a^2 + \frac{m-1}{4}\left(1 - \frac{1}{N}\right)^2}.$$

Since a is arbitrarily close to 1 for sufficiently large N, this proves that

$$K_m(0, 0) \geqslant \sqrt{1 + \frac{m-1}{4}} = \frac{1}{2}\sqrt{m+3}.$$

Now let $e_0, e_1, \ldots, e_m$ be mutually orthogonal unit vectors in $(m+1)-$ dimensional space and let N, a, v_i be as defined in the preceding paragraph. Set $k = m + 4N - 1$, let M be a positive integer and define vectors as follows:

$$u_1 = e_0$$

$$w_{ri} = \sqrt{1 - \frac{1}{k^2M^2}}\, v_i - \frac{e_0}{kM} \text{ for } 1 \leqslant i < m + 4N,\ 1 \leqslant r \leqslant M.$$

Then $|u_1| = 1$, $|w_{ri}| = 1$ for all r, i and $u_1 + \sum_{r,i} w_{ri} = 0$. Yet for every arrangement of the vectors w_{ri} preceded by u_1 as first vector, one of the partial sums which contains less than k terms has length not less than the smaller of

$$\sqrt{2N} \text{ and } \sqrt{\left(1 - \frac{1}{M}\right)^2 + \left(1 - \frac{1}{k^2M^2}\right)\left(a^2 + \frac{m-1}{4}\left(1 - \frac{1}{N}\right)^2\right)}.$$

Since N, M can be taken arbitrarily large and a arbitrarily close to 1 it follows that $K_{m+1}(1, 0) \geqslant \sqrt{1 + 1 + \frac{m-1}{4}} = \frac{1}{2}\sqrt{m+7}$. This means that $K_m(1, 0) \geqslant \frac{1}{2}\sqrt{m+6}$ for all $m \geqslant 2$. In particular, $K_2(1, 0) \geqslant \frac{1}{2}\sqrt{8} = \sqrt{2}$.

REFERENCES

[1] E. Steinitz. *Bedingt konvergente Reihen und konvexe Systeme,* Journal f. Math., vol. 143 (1913), 128-175.

[2] W. Gross. *Bedingt konvergente Reihen,* Monatshefte F. Math U. Physik, vol. 28 (1917), 221-237.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLIV : SERIES III : JUNE, 1950
SECTION THREE

Examples of Integrals that are Discontinuous in Sets of Positive Measure*

BY H. W. ELLIS

Presented by R L. JEFFERY, F R.S.C.

1. Introduction. Definitions of integration in which the continuity of the indefinite integral is replaced by the weaker properties of approximate continuity or of mean continuity have been studied by several authors.[1] We shall restrict our discussion to the four definitions outlined below.

DEFINITION. *The function $F(x)$ is Cesàro or M_1-continuous at the point $x = x_0$ if $F(x)$ is integrable in the special Denjoy sense in the neighbourhood of x_0 and if*

$$\lim_{h \to 0} \frac{1}{h} \int_{x_0}^{x_0+h} F(t)dt = F(x_0).$$

DEFINITION. *The upper and lower two-sided Cesàro derivatives of $F(x)$ at x_0, $\overline{CD}F$ and $\underline{CD}F$ are defined respectively as the upper and lower limits as $h \to 0$ of*

$$\frac{2}{h^2}\int_{x_0}^{x_0+h} \{F(t) - F(x)\}\, dx,$$

where it is implied that $F(t)$ is integrable in the special Denjoy sense in the neighbourhood of x_0. If these two limits are equal their common value is CDF, the Cesàro derivative of $F(x)$ at x_0.

Definitions of Burkill's Cesàro-Perron (CP) [2] and approximate Perron (*AP*) [1] integrals are then obtained from the definition of the Perron integral by replacing ordinary continuity and derivatives in the definitions of major and minor functions by Cesàro continuity and Cesàro derivatives for the *CP*-integral and by approximate continuity and approximate derivatives [5, p. 220] for the *AP*-integral.

We next give descriptive definitions of Ridder's β-integral [4] and of the generalized mean integral of order one (GM_1-integral) [3].

---

* This study was completed while the author was holder of a fellowship at the Summer Research Institute of the Canadian Mathematical Congress.

[1] For a partial bibliography see [3].

The function $f(x)$ is β-integrable (GM_1-integrable on $[a, b]$ if there exists an approximately continuous (M_1-continuous) function $F(x)$ such that $[a, b]$ can be covered by a denumerable sequence of closed sets over each of which $F(x)$ is absolutely continuous and such that $ADF(x) = f(x)$ almost everywhere on $[a, b]$. The function $F(x)$ is called an indefinite β-integral (GM_1-integral) of $f(x)$ and the definite integral of f on $[a, b]$ is $F(b) - F(a)$.

Examples so far given of functions that are indefinite integrals in the above senses have single points of discontinuity and no consideration has been given to the possible distribution of the points of discontinuity. In §2 we define a class of bounded functions on [0, 1], discontinuous in sets of points of positive measure, that are indefinite AP-, β-, and GM_1-integrals on [0, 1]. In §3 we show that, although all of these functions are Cesàro continuous at every point of [0, 1] some are indefinite CP-integrals while others are not. Those that fail to be CP-integrals do so because the points at which their Cesàro derivatives fail to exist or fail to be finite have positive measure. In particular this last class gives examples of functions that are GM_1-integrable without being CP-integrable.

The following simple theorem and corollary are established in §2.

THEOREM. *If $F(x)$ is approximately continuous at x_0, measurable and bounded in the neighbourhood of x_0, then $F(x)$ is both Cesàro and M_1-continuous at x_0.*

COROLLARY. *If the bounded function $F(x)$ is an indefinite β-integral of $f(x)$ on $[a, b]$ then $f(x)$ is GM_1-integrable on $[a, b]$ and $F(x)$ is also an indefinite GM_1-integral of f on $[a, b]$.*

2. Let $\{\lambda_n\}$ be a sequence of positive numbers with $\sum_{n=1}^{\infty} \lambda_n < 1$. Let I_{11} be an open interval with centre coinciding with the centre of the interval [0, 1] and length λ_1 and let J_{11}, J_{12} denote the closed intervals complementary to I_{11} on [0, 1]. Proceeding inductively we define, for $n = 2, 3, \ldots$, the open intervals I_{ni}, $i = 1, 2, \ldots, 2^{n-1}$ each of length $\lambda_n/2^{n-1}$ and with centres coinciding with the centers of the closed intervals $J_{n-1,i}$, $i = 1, 2, \ldots 2^{n-1}$ and define J_{ni}, $i = 1, 2, \ldots, 2^n$ as the equal closed intervals complementary to $\sum_{r=1}^{n} \sum_{j=1}^{2^{r-1}} I_{rj}$. Let E denote the perfect set of points of [0, 1] complementary to $\sum_{i=1}^{\infty} \sum_{j=1}^{2^{r-1}} I_{ij}$. Then $mE = 1 - \sum_{n=1}^{\infty} \lambda_n > 0$ since $\sum_{1}^{\infty} \lambda_n < 1$. Simple computations show

that the length of each interval J_{ni} is $\frac{1}{2^n}\left(1 - \sum_{r=1}^{n} \lambda_r\right)$ and the measure of the points of $CE.J_{ni}$ is $\frac{1}{2^n} \sum_{r=n+1}^{\infty} \lambda_r$.

Let $K > 1$ be fixed and let $\{\mu_n\}$ be a sequence of positive numbers each less than unity and with $\lim \mu_n = 0$. On each of the intervals $I_{ni} = (c_{ni}, d_{ni})$ for all n and i define a set E_{ni} of non-overlapping intervals such that: (i) $mE_{ni} = \mu_n . |I_{ni}|$, (ii) c_{ni}, d_{ni} are limit points of E_{ni} with the right density of E_{ni} zero at c_{ni}, the left density of E_{ni} zero at d_{ni}, (iii) there are only a finite number of intervals of E_{ni} on any closed interval contained in I_{ni} and

$$\text{(iv)} \qquad \max_{c_{ni} \leqslant \xi \leqslant d_{ni}} \frac{mE_{ni}(c_{ni}, \xi)}{\xi - c_{ni}} < K\mu_n,$$

$$\max_{c_{ni} \leqslant \xi \leqslant d_{ni}} \frac{mE_{ni}(\xi, d_{ni})}{d_{ni} - \xi} < K\mu_n,$$

for all values of n and i.

Set $E_* = \sum_{i,n} E_{ni}$ and define $F(x) = 0$ at the points of $CE_*(0, 1)$. On each of the intervals of each set E_{ni} as base construct an isosceles triangle with altitude unity. For x a point of the base of any of these triangles define $F(x)$ as the corresponding ordinate to the side. Define $f(x) = ADF(x)$ at the points where the approximate derivative exists and let $f(x) = 0$ elsewhere on $[0, 1]$. We shall show that $f(x)$ is AP-, β-, and GM_1-integrable on $[0, 1]$ and that $F(x)$ is an indefinite AP-, β-, and GM_1-integral of $f(x)$ on $[0, 1]$. Furthermore every neighbourhood of every point x of E contains points at which $F(x)$ is zero and points at which $F(x)$ is unity so that $F(x)$ is discontinuous at every point of E where E has positive measure.

We shall show that the density of E_* (i.e. of the points at which $F(x) \neq 0$) is zero at every point of E, from which it follows that $F(x)$ is approximately continuous and $ADF(x) = 0$ at every point of E. For each of the equal intervals $J_{ni} = [a_{ni}, b_{ni}]$ and all n let

$$r_n = \max_{a_{ni} \leqslant \xi \leqslant b_{ni}} \frac{mCE(a_{ni}, \xi)}{\xi - a_{ni}} = \max_{a_{ni} \leqslant \xi \leqslant b_{ni}} \frac{mCE(\xi, b_{ni})}{b_{ni} - \xi}.$$

We shall show that $\lim r_n = 0$. Consider the point $x = 0$. This point is the left end point of the intervals J_{n1}, $n = 1, 2, \ldots$. For $b_{n1} \leqslant \xi \leqslant b_{n-1,1}$

$$(1) \qquad \frac{mCE(0,\xi)}{\xi} < \frac{mCE(0,b_{n-1,1})}{b_{n-1,1}} = \frac{2\sum_{n}^{\infty}\lambda_r}{1-\sum_{r=1}^{n}\lambda_r}$$

which approaches zero as $n \to \infty$, since $\sum_{1}^{n}\lambda_r$ approaches a limit less than unity. It follows that the right density of CE is zero at $x = 0$. Since each interval with first subscript n is identical with J_{n1} and the points of CE are distributed on the interval symmetrically with respect to its midpoint, (1) implies that $r_n \to 0$ as $n \to \infty$. For each interval $I_{ni} = (c_{ni}, d_{ni})$ there are relations similar to (1) for points of CE on the left at c_{ni} and on the right at d_{ni} and, remembering (ii) above, it follows that the density of E_* is zero at c_{ni} and at d_{ni} for all values of n and i.

Consider any point $x \in E$ that is not an end point of one of the intervals I_{ni}. There then exists a sequence I_{nq} of the intervals I_{ni} approaching x with x contained in the interval $J_{nq} = [a_{nq}, b_{nq}]$ contiguous to I_{nq} on the right. Let $J'_{nq} = [a'_{nq}, b'_{nq}]$ be the corresponding closed intervals contiguous to I_{nq} on the left. For $a'_{nq} \leqslant \xi \leqslant b'_{nq}$,

$$\frac{mE_*(\xi,x)}{x-\xi} < \frac{mCE(\xi,b'_{nq})}{b'_{nq}-\xi} + \frac{mE_{nq}}{|I_{nq}|} + \frac{mCE(a_{nq},x)}{x-a_{nq}} < 2r_{nq} + \mu_{nq}.$$

For $\xi \in I_{nq}$,

$$\frac{mE_*(\xi,x)}{x-\xi} < \frac{mE_{nq}(\xi,a_{nq})}{a_{nq}-\xi} + r_{nq} < K\mu_{nq} + r_{nq}.$$

In both cases the right side is arbitrarily small for n_q sufficiently large. We conclude that the density of E^* is zero on the left at x. It may be shown in the same way that E^* has right density zero at x. We have therefore shown that E^* has density zero at each point of E. From (iii) in the definition of the sets E_{ni} it follows that $F(x)$ is continuous at each point of CE and has an ordinary derivative at all except a denumerable set of the points of CE. The function $F(x)$ is approximately continuous at every point of [0, 1] and $ADF(x)$ exists and equals $f(x)$ at all points of [0, 1] except those of a denumerable set. Since $F(0) = 0$, $F(x)$ serves as both a major and minor function to $f(x)$ on [0, 1] in the sense of the AP-integral. It follows that $f(x)$ is AP- and β-integrable to $F(x)$ on [0, 1] [4, p. 154].

Proof of the theorem. There is no loss of generality in considering one sided continuity at $x = a$. For δ sufficiently small $|F(x)| < M$ on $[a, a+\delta]$. Let $\epsilon > 0$ be given and for $a < x < a + \delta$, let $S(a, x)$ be the set of points ξ of (a, x) with $|F(\xi) - F(a)| < \epsilon/2$. The function $F(x)$ is bounded and therefore Lebesgue integrable on $[a, a + \delta]$ and

$$\left|\frac{1}{x-a}\int_a^x \{F(t) - F(a)\}dt\right| = \left|\frac{1}{x-a}\int_{S(a,x)} \{F(t) - F(x)\}\, dt\right.$$

$$+\frac{1}{x-a}\int_{CS(a,x)} \{F(t) - F(x)\}\, dt$$

$$< \frac{mS(a,x)}{x-a}\,\epsilon + \frac{mCS(a,x)}{x-a}\,.\,2M$$

$$< \epsilon,$$

for all x sufficiently near a since the density of S is unity and the density of CS zero on the right at $x = a$. We conclude that $F(x)$ is both Cesàro and M_1-continuous on the right at a. That the converse of the theorem is not true is shown by the function $F(x) = \sin \dfrac{1}{x}$, $x \neq 0$; $F(0) = 0$ which is Cesàro and M_1-continuous at the origin without being approximately continuous there. The corollary stated in the introduction is an immediate consequence of the definitions of β- and GM_1-integrability and shows that $f(x)$ is GM_1-integrable to $F(x)$ on $[0, 1]$.

3. The theorem shows that the function $F(x)$ corresponding to any pair of sequences $\{\lambda_n\}$ and $\{\mu_n\}$ is Cesàro continuous on $[0, 1]$. We next show that if these sequences are such that $2^n \sum\limits_{r=n+1}^{\infty} \mu_r\lambda_r \to \infty$ as $n \to \infty$ then $f(x)$ cannot be CP-integrable on $[0, 1]$.

If $F(x)$ is an indefinite CP-integral on $[0, 1]$ then $CDF(x)$ must exist and be finite at almost all points of $[0, 1]$ [2]. If $F(x)$ is integrable in the special Denjoy sense, $\{J_k = [a_k, b_k]\}$ a sequence of intervals containing x and converging to x and if $F(a_k) = F(b_k) = 0$ for all values of k, the relation

$$\frac{2}{(b_k - a_k)^2}\int_{a_k}^{b_k} F(t)dt = \left(\frac{x - a_k}{b_k - a_k}\right)^2 \frac{2}{(x - a_k)^2}\int_{a_k}^{b_k} F(t)dt$$

$$+\left(\frac{b_k - x}{b_k - a_k}\right)^2 \frac{2}{(b_k - x)^2}\int_{a_k}^{b_k} F(t)dt$$

shows that when $CDF(x)$ exists and is finite the limit as $k \to \infty$ of the expression on the left exists and is finite. Conversely if the upper limit as $k \to \infty$ of the expression on the left is infinite $CDF(x)$ cannot exist and be finite at x.

The points of J_{ni} at which $F(x) \neq 0$ have measure $\frac{1}{2^n} \sum_{r=n+1}^{\infty} \mu_r \lambda_r$ whence $\int_{a_{ni}}^{b_{ni}} F(t)dt = \frac{1}{2^{n-1}} \sum_{n+1}^{\infty} \mu_r \lambda_r$. Then each point $x \in E$ that is not an end point of an interval I_{ni} is contained in an infinite sequence of intervals J_{ni} converging to x and

$$\frac{2}{(b_{ni} - a_{ni})^2} \int_{a_{ni}}^{b_{ni}} F(t)dt = \frac{2^n \sum_{n+1}^{\infty} \mu_r \lambda_r}{[1 - \sum_{1}^{n} \lambda_r]^2}$$

which increases without bound provided $2^n \sum_{n+1}^{\infty} \mu_r \lambda_r \to \infty$ as $n \to \infty$. With this supposition $CDF(x)$ cannot exist and be finite at more than a denumerable set of the points of E and therefore fails to exist at a set of positive measure. The function $f(x)$ is then not CP-integrable. For example if $\lambda_r = \frac{k}{(r+1)(r+2)}$, $0 < k < 1$, $\mu_r = 1/r$, the corresponding function $F(x)$ is Cesàro continuous at each point of $[0, 1]$ but $f(x)$ is not CP-integrable.

Finally a subclass of the functions in §2 that are indefinite CP-integrals may be obtained by imposing the added restrictions on the sequences $\{\lambda_n\}$, $\{\mu_n\}$ that $\lim_{n \to \infty} \frac{\mu_n}{|I_{ni}|} = \lim \frac{2^{n-1}\mu_n}{\lambda_n} = 0$, $2^n \sum_{n+1}^{\infty} \mu_r \lambda_r \to 0$ as $n \to \infty$, where now the intervals of the sets E_{ni} must be so distributed that (iv) can be replaced by (iv′)

$$\max_{c_{ni} \leqslant \xi \leqslant d_{ni}} \frac{mE_{ni}(c_{ni}, \xi)}{(\xi - c_{ni})^2} < K\mu_n, \qquad \max_{c_{ni} \leqslant \xi \leqslant d_{ni}} \frac{mE_{ni}(\xi, d_{ni})}{(d_{ni} - \xi)^2} < K\mu_n$$

where $K > 2$ is fixed for all n.

REFERENCES

[1] J. C. Burkill, *The approximately continuous Perron integral*, Math Zeitschrift, vol. 34 (1931), 270-8.

[2] ——— *The Cesàro-Perron integral*, Proc. Lond. Math. Soc., (2) vol. 34 (1932), 314-22.

[3] H. W. Ellis, *Mean-continuous integrals*, Can. J. Math., vol. I (1949), 113-24.

[4] J. Ridder, *Über die gegenseitigen Beziehungen verschiedener approximativ stetiger Denjoy-Perron Integrale*, Fundamenta Mathematicae, vol. XXII (1934), 136-62.

[5] S. Saks, *Theory of the integral* (Warsaw, 1937).

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLIV : SERIES III : JUNE, 1950
SECTION THREE

Limit Points of Riemann Sums*

By R. L. JEFFERY, F R S C.

THIS note is a supplement to a paper by C. N. Rowse and the present writer [1]. It was pointed out to us by I. Halperin that we were in error in assuming the mean convergence of the sequences of functions ψ_n' which are used in the proof of Theorem II. What is essential for the proof, as we have given it, is to know that every subsequence of a sequence ψ_n' [j fixed], contains a convergent subsequence. However, it is easy to construct sequences of such functions which contain no convergent subsequences.

Divide [0, 1] in halves, then in quarters, and so on. At the nth stage of subdivision the number of subintervals is 2^n. Denote these by (x_{i-1}, x_i) $i = 1, 2 \ldots 2^n$. Define $\psi_n(x)$ to be 1 on the intervals (x_{i-1}, x_i) with i odd, and zero elsewhere.

We now verify that the sequence ψ_n contains no subsequence which converges on a set of positive measure. If some subsequence ψ_{n_i} were convergent on a set of positive measure, then its limit-function would take either the value 0 or the value 1 on a set of positive measure.

Let E_n be the set on which $\psi_n(x) = 1$. Obviously $|E_n| = \frac{1}{2}$, and

$$|E_n E_{n+1} \ldots E_{n+l-1}| = \frac{1}{2^{-l}}. \tag{1}$$

Hence, if E^*_p denotes the set on which $\psi_n(x) = 1$ for all $n > p$, then (1) implies $|E^*_p| = 0$. This contradicts the supposition that $\psi_{n_i} \rightarrow 1$ on a set of positive measure. But the same argument applies to the limit-value zero and this is obviously contradictory.

Thus the proof of Theorem II which we have given holds only in special cases. We have, nevertheless, been able to arrive at the proof of this theorem by a different method. In this case, as in the case of the paper we are supplementing, the incidental results are of more interest than the main theorem, for, as noted in the former paper, this has been proved in a different way by Hartman.

We first prove two results concerning functions of a real variable.

---

* This study was completed at the Summer Research Institute of the Canadian Mathematical Congress.

(i) *Let the function* $g(x)$ *be summable on* $[a, b]$. *Let* (x_{i-1}, x_i) *be a subdivision of* $[a, b]$, *and let* x^0_i *be a point with* $x_{i-1} < x^0_i < x_i$, *and such that* $x^0_i - x_{i-1} = \lambda(x_i - x_{i-1})[0 < \lambda < 1]$. *Then*

$$\left| \Sigma \int_{x_{i-1}}^{x^0_i} gdx - \lambda \int_a^b gdx \right| \rightarrow 0,$$

as Max $(x_i - x_{i-1}) \rightarrow 0$.

Since g is measurable there is a closed set C on $[a, b]$ such that g is continuous on C, and such that both $b - a - |C|$ and

$$\left| \int_a^b gdx - \int_C gdx \right| \tag{2}$$

are arbitrarily near zero. It is known [1] that if (x_{i-1}, x_i) is a subdivision of $[a, b]$ with Max $(x_i - x_{i-1})$ sufficiently small, and if we denote by (x_k, x'_k) the intervals of this subdivision which contain some ξ_k of C as interior point, then

$$| \Sigma\, g(\xi_k)(x'_k - x_k) - \int gdx | \tag{3}$$

is arbitrarily near zero. Now let $\epsilon > 0$ be given and take Max (x_{i-1}, x_i) so small that if x and x' are two points of C on any interval (x_{i-1}, x_i), then $|g(x) - g(x')| < \epsilon$. Take ξ^0_k and ξ_k to be points of C on (x_k, x^0_k) and (x_l, x'_k) respectively, then

$$\begin{aligned} g(\xi^0_k)(x^0_k - x_k) &= [g(\xi_k) + \epsilon_k](x^0_k - x_k) \\ &= \lambda \Sigma\, [g(\xi_k) + \epsilon_k][x'_k - x_k] \\ &= \lambda \Sigma\, g(\xi_k)(x'_k - x_k) + \lambda \Sigma \epsilon_k (x'_k - x_k) \end{aligned}$$

where $\epsilon_l < \epsilon$. Hence

$$|\Sigma\, g(\xi^0_k)(x^0_k - x_k) - \lambda \Sigma\, g(\xi_k)(x'_k - x_k)| < \epsilon(b - a). \tag{4}$$

Now for x on $C(x_k, x^0_k)$, we have $|g(x) - g(\xi^0_k)| < \epsilon$. Hence

$$\int_{C(x_k,\, x^0_k)} gdx - \Sigma\, g(\xi^0_l)\, ||C.(x_k, x^0_k)| < \epsilon(x^0_k - x_k). \tag{5}$$

Also

$$\left| \int_{x_k}^{x^0_k} g(x)dx - \Sigma \int_{C(x_k,\, x^0_k)} gdx \right| < \Sigma \int_{e_k} |g| dx \tag{6}$$

where $e_k = (x_k, x^0_k) - C(x_k, x^0_k)$. Since g is summable, the left side of this inequality is as small as we wish provided $|C|$ is sufficiently close to $b - a$. Again

$$g(\xi^0_k)(x^0_k - x_k) - g(\xi^0_j)|C(x_k, x^0_k)| < M\sum e_r, \tag{7}$$

where $M = \sup |g(x)|$, x on C. Once C has been fixed, and consequently M, $\sum|e_k|$ can be made arbitrarily small by taking Max $(x_i - x_{i-1})$ sufficiently small. Since in the foregoing ϵ is arbitrary and independent of the choice of C relations (2)–(7) combine to give the desired result.

(ii) *Let the functions $G(x)$ be absolutely continuous on $[a, b]$. Let (x_{i-1}, x_i) be a subdivision of $[a, b]$, x^0_i a point with $x_{i-1} < x^0_i < x_i$ and such that $x^0_i - x_{i-1} = \lambda(x_i - x_{i-1})$, $0 < \lambda < 1$. If Max $(x_i - x_{i-1})$ is sufficiently small, then*

$$\left| \sum \{G(x^0_i) - G(x_i)\} - \lambda\{G(b) - G(a)\} \right|$$

is arbitrarily near zero.

Since G is absolutely continuous

$$G(x^0_i) - G(x_i) = \int^{x^0_i} G'dx,$$

$$G(b) - G(a) = \int_a^b G'dx,$$

and (ii) now follows from (i).

We note that the absolute continuity of G is essential for the validity of (ii). This we show by an example. Let E be the Cantor ternary set on $[0, 1]$ and let ϕ be the corresponding Cantor function. Let u_i be the intervals which contain E at the nth stage of the deletion process which gives rise to E. Let $u_i = (x_i, x^0_i)$. There is to the right of each u_i (except the interval with right-hand end-point equal to unity) a deleted interval or a part of a deleted interval (x^0_i, x'_i) with $x'_i - x^0_i = x^0_i - x_i = \lambda(x'_i - x_i)[\lambda = \frac{1}{2}]$. As n increases $|u_i|$ and consequently Max $(x'_i - x_i) \to 0$. It now follows that if (ii) holds for the function ϕ then, for ϵ arbitrary, and all n sufficiently great,

$$\sum \{ \phi(x^0_i) - \phi(x_i) \} < \tfrac{1}{2} \{ \phi(1) - \phi(0) \} + \epsilon < \tfrac{1}{2} + \epsilon.$$

But $\sum \{ \phi(x^0_i) - \phi(x_i) \} \to 1$ as $n \to \infty$. This is a contradiction and the desired result is established.

We are now ready to proceed with the main problem. Let $\Sigma_n = \sum f(\xi_{ni})(x_{ni} - x_{n,i-1})$ be a sequence of Riemann sums defined on $[a, b]$, where $x_{n,i-1} \leqslant \xi_{ni} \leqslant x_{ni}$, and such that

$$\lim_{n\to\infty} \Sigma_n = U, \quad \operatorname*{Max}_i (x_{ni} - x_{n,i-1}) \to 0.$$

It is no restriction to assume $0 < f(\xi_{ni}) < M$; we also note that the definition of the function f can be restricted to the points ξ_{ni}.

We now define two functions $f_n(x)$ and $\psi_n(x)$ by

$$f_n(0) = 0, \qquad f_n(x) = f(\xi_{ni}), \qquad x_{n,\,i-1} < x \leqslant x_{ni},$$

$$\psi_n(x) = \int_a f_n(x)dx.$$

Obviously

$$\psi_n(a) = 0, \qquad \psi_n(b) = \Sigma_n, \qquad \lim_{n \to \infty} \psi_n(b) = U.$$

It is also obvious that $\psi_n(x)$ is a continuous, non-decreasing function, which satisfies the Lipschitz condition

$$(8) \qquad \psi_n(x') - \psi_n(x) < M(x' - x), \qquad (x' > x).$$

We now prove

(iii) *There is a convergent subsequence $\psi_{n_j}(x)$ of the sequence $\psi_n(x)$ which tends to an absolutely continuous function $\psi(x)$.*

By Helly's theorem [2], the sequence ψ_n contains a convergent subsequence, ψ_{n_j}. Furthermore, since relation (8) holds, uniformly in n, it follows that $\psi(x') - \psi(x) \leqslant M(x' - x)$, which implies the absolute continuity of $\psi(x)$. This completes the proof of (iii).

Now let $\Sigma f(\xi'_{ni})(x'_{ni} - x'_{n,\,i-1})$ be another sequence of Riemann sums with limit L. This sequence gives rise to a sequence $\phi_n(x)$ which tends to an absolutely continuous function $\phi(x)$ with $\phi(a) = 0$, $\phi(b) = L$. Given λ, $[0 < \lambda < 1]$, consider

$$(9) \qquad \Sigma \{\psi(x^0_i) - \psi(x_i)\} - \lambda\{\psi(b) - \psi(a)\}$$

$$(10) \qquad \Sigma \{\phi(x_i) - \phi(x^0_i)\} - (1 - \lambda)\{\phi(b) - \phi(a)\},$$

where $\quad x_{i-1} < x^0_i < x_i, \quad x^0_i - x_{i-1} = \lambda(x_i - x_{i-1}),$

$$\psi(b) - \psi(a) = \psi(b) = U, \qquad \phi(b) - \phi(a) = \varphi(b) = L;$$

making use of (ii), (9) and (10) can be made as small as we please by taking Max $(x_i - x_{i-1})$ sufficiently small. Consider the subsequences of the Riemann sums which correspond to the sequences ψ_n and ϕ_n respectively. Denote these Riemann sums by

$$(11) \qquad \Sigma f(\xi_j)(t_j - t_{j-1}), \qquad \Sigma f(\xi^0_j)(t'_j - t'_{j-1})$$

respectively. Let (t_{ip}, t^0_{ip}) be intervals of the set (t^0_{iq}, t_i) which are contained by the interval (x_{i-1}, x^0_i), and (t^0_{iq}, t_{iq}) intervals of the set (t'_{j-1}, t'_j) which are contained by the interval (x^0_i, x_i). Then, for

$$d = \text{Max}\, [t_i - t_{i-1}), \quad (t'_j - t_{j-1})] \tag{12}$$

sufficiently small, we have

$$f(\xi_{ip})(t^0_{ip} - t_{ip}) - \Sigma \left\{ \psi(x^0_i) - \psi(x_{i-1}) \right\}, \tag{13}$$

$$f(\xi_{iq})(t_{iq} - t^0_{iq}) - \Sigma \left\{ \phi(x_i) - \phi(x^0_i) \right\}, \tag{14}$$

each arbitrarily near zero. Let (x_r, x'_r) be the intervals on $[a, b]$ complementary to the sets (t_{ip}, t^0_{ip}), (t^0_{iq}, t_{iq}) ξ_r any point on (x_r, x'_r) for which $f(x)$ is defined. Such points exist on every interval because each of the sets ξ_{ni}, ξ'_{ni} is everywhere dense. We then have

$$\left| f(\xi_r)(x'_r - x_r) \right| < M \Sigma \, (x'_r - x_r) \tag{15}$$

where $M = \sup f(x)$, x on the sets ξ_{ni}, ξ'_{ni}. Furthermore, $\Sigma \, (x'_r - x_r)$ is arbitrarily small if only d, in (12), is sufficiently small. Now, making use of (9), (10), (11), (13), (14) we have

$$\Sigma_i \Sigma_p f(\xi_{ip})(t^0_{ip} - t_{ip}) + \Sigma_i \Sigma_q f(\xi_{iq})(t_{iq} - t^0_{iq}) + \Sigma_r f(\xi_r)(x'_r - x_r) \tag{16}$$

arbitrarily near to

$$\lambda U + (1 - \lambda)L = L + \lambda(U - L).$$

Since λ is any number between zero and unity, it follows that given the original Riemann sums, it is possible to determine from them a Riemann sum which is arbitrarily near a predetermined point on the interval defined by the numbers L, U.

What we have done appears to be a long way around to a simple result. However, the methods we have used now combine to give us much more. Let

$$F(x) = [f_1(x), \ f_2(x), \ldots, f_k(x)]$$

be a vector function in k-dimensional Euclidean space, where $f_j(x)$ is defined on $[a, b]$ and $0 \leqslant f_j(x) < M$, $j = 1, 2, \ldots, k$. Suppose there are sequences of subdivisions (x_{i-1}, x_i), (x'_{i-1}, x'_i) of $[a, b]$ which give rise to vectors

$$[f_1(\xi_i) \ (x_i - x_{i-1}), \ldots, f_k(\xi_i) \ (x_i - x_{i-1})],$$
$$[f_1(\xi'_i) \ (x'_i - x'_{i-1}), \ldots, f_k(\xi'_i)(x'_i - x'_{i-1})],$$

which tend respectively to the vectors

$$[U_1, U_2, \ldots, U_k], \quad [L_1, L_2, \ldots, L_k].$$

The functions f_j being given, consider the sequences ψ_{1n}, ϕ_{1n} tending, respectively, to the absolutely continuous functions ψ_1, ϕ_1.

Let

(17) $$\Sigma f_1(\xi_i)(t_i - t_{i-1}), \quad \Sigma f_1(\xi'_i)(t'_i - t'_{i-1}),$$

be the associated sequences of Riemann sums [corresponding to those in (11)]. Now, the sequences of subdivisions (t_{i-1}, t_i) and (t'_{i-1}, t'_i) correspond to subsequences

(18) $$\Sigma f_2(\xi_i)(t_i - t_{i-1}), \quad \Sigma f_2(\xi'_i)(t'_i - t'_{i-1})$$

and, from the sequences in (18) it is possible to determine the sequences ψ_{2n}, ϕ_{2n} tending to the absolutely continuous functions ψ_2, ϕ_2 and the subsequences of the sequences in (18) which correspond to those in (17),

(19) $$\Sigma f_2(\xi_i)(s_i - s_{i-1}), \quad \Sigma f_2(\xi'_i)(s'_i - s'_{i-1}).$$

We can now start with the sequences in (19), work as we did with those in (17), to arrive at the corresponding functions and sequences for the function f_3. Continuing, we arrive at the sequences of functions ψ_{kn}, ϕ_{kn} tending to the absolutely continuous functions ψ_k, ϕ_k, and the related sequences of Riemann sums

$$\Sigma f_k(\xi_i)(u_i - u_{i-1}), \quad \Sigma f_k(\xi'_i)(u_i - u'_{i-1}).$$

Furthermore, the sequences of subdivisions (u_{i-1}, u_i), (u'_{i-1}, u'_i) determine sequences of Riemann sums

$$\Sigma f_j(\xi_j)(u_i - u_{i-1}), \quad \Sigma f_j(\xi'_j)(u_i - u_{i-1}),$$

$j = 1, 2, \dots, k - 1$, which are related to the functions ψ_j, ϕ_j,
$j = 1, 2, \dots, k - 1$ as the sequences in (11) are related to the functions ψ_n, ϕ_n. Consequently, for Max $(x_i - x_{i-1})$ sufficiently small, it is possible to determine k Riemann sums similar to (6) with j, replacing f and the sums arbitrarily near the respective limits

$$L_1 + \lambda(U_1 - L_1), \quad L_2 + \lambda(U_2 - L_2), \dots, L_k + \lambda(U_k - L_k).$$

Since λ is any number between zero and unity it follows that there are Riemann sums arbitrarily near to every point on the line joining the points

$$P_1 = [U_1, \dots, U_k], \quad P_2 = [L_1, \dots, L_k].$$

Thus if P_1 and P_2 are any two points on the set of limit points of the vectors

$$f_1(\xi_i)(x_i - x_{i-1}), \dots, f_k(\xi_i)(x_i - x_{i-1})$$

with Max $(x_i - x_{i-1}) \to 0$ then every point on the line joining P_1 and P_2 is a limit point, hence the set of limit points of the set of vectors is a convex closed set. This completes the proof of Theorem II of [1].

A question which was raised in the proof of Theorem I [1] is the following: If the Riemann sums $\Sigma f(\xi_i)(x_i - x_{i-1})$ tend to a limit L as Max $(x_i - x_{i-1})$ tends to zero, then the totality of points ξ_i constitute an everywhere dense set E on $[a, b]$. Is it in general possible to select a subset E' of E which is everywhere dense on $[a, b]$, and which is such that if (x_{i-1}, x_i) is any sequence of subdivisions of $[a, b]$ and ξ'_i any point of E' on $[x_{i-1}, x_i]$ then $\Sigma f(\xi'_i)(x_i - x_{i-1})$ tends to L? We construct an example which answers this question in the negative. Let G be a Cantor set on $[0, 1]$ with $|G| > 0$. Let $f(x) = 1$ when $x \epsilon G$, $f(x) = 0$ otherwise. Let (x_{i-1}, x_i) be any sequence of subdivisions of $[a, b]$ with Max $(x_i - x_{i-1}) \to 0$, let (x_k, x'_k) be the intervals of this set which contain points of G, and let ξ_k be any point of G on (x_k, x'_k). For the remaining intervals (x_r, x'_r) of the set (x_{i-1}, x_i) take ξ_r to be any point on the interval. If ξ_r is not a point of G then $f(\xi_r) = 0$. Also

$$\Sigma_k f(\xi_k)(x'_k - x_k) + \Sigma_r f(\xi_r)(x'_r - x_r) \to S\, f dx = |G|.$$

Now any everywhere dense set E' on $[0, 1]$, is everywhere dense on the complement $\tilde{G}$ of G; moreover, $\tilde{G}$ is everywhere dense $[0, 1]$. Hence, regardless of the nature of the sequence of subdivisions (x_{i-1}, x_i) of $[0, 1]$, it is possible to select a point ξ'_i of E' so that

$$\Sigma f(\xi'_i)(x_i - x_{i-1}) = 0.$$

REFERENCES

[1] R. L. Jeffery and C. N. Rowse, *The limit points of Riemann sums*, Trans. Roy. Soc. Can., vol. 43 (1949), 21-26.

[2] E. Helly, *Über lineare Funktionaloperationen*. Sitzungsberichte der Wissenschaften in Wien, vol. 121, IIa (1912), 283.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLIV : SERIES III : JUNE, 1950
SECTION THREE

Correlation of Auroral Observations in the Northern and Southern Hemispheres

By D. E. LITTLE and G. M. SHRUM, F.R.S.C

THE present paper represents the results of a survey of literature and unpublished records of observations of the Aurora Borealis and the Aurora Australis. The object has been to determine whether or not a positive correlation exists between the occurrence of the two auroral phenomena in the Northern and Southern Hemispheres. In making this study a number of unforeseen difficulties have been encountered. The main problem has been the lack of suitable observations. In the Southern Hemisphere it is only recently that any complete details of the occurrences of aurorae have been recorded, whereas in the Northern Hemisphere fairly complete records of occurrences are available for the latter half of the nineteenth century but relatively few such observations are being recorded at the present time. Unfortunately, too, published observations of aurorae are usually incomplete in any one library or publication. In order to overcome this, a literature survey was made, partly by means of inter-library loans of the following types of journals and periodicals in both scientific and popular form: General Science, Astronomy, Astrophysics, Geophysics, Terrestrial Magnetism, Meteorology, Geography, and Physics. Also, accounts of expeditions and voyages in both hemispheres were examined.

Literature and Data. The generally accepted theory of aurorae is that attributed to Stormer [1], [4], [16], or a slight modification of it. Stormer [13], *circa* 1930, assumed that charged particles are emitted from the sun, that these particles have velocities approaching that of light, and that they are acted on only by forces due to the earth's magnetic field, which he considered to be represented by a magnetic dipole at the earth's centre. From equations of motion, he found trajectories of these particles by numerical integration, which agreed with the empirical results Birkeland, 1896, had found before for cathode rays impinging on a magnetically charged sphere or "terella." The lower latitude of the auroral zone was fixed by a parameter of mathematical origin (for other values of this parameter, the particles miss the earth). The upper latitude was determined by the permitted incident angles for the particles entering the earth's atmosphere. In order to enlarge this auroral zone to compare with the zone of maximum auroral activity observed on the earth, Stormer assumed ring currents about the earth, far above the equator. Brüche [2] did a subsequent series of experiments which illustrated Stormer's theory.

Many investigators consider auroral occurrences to be associated with magnetic storms [4], [10], [11], [12], [17]. If this is true it does indicate that a correlation between auroral observations in the two hemispheres might be expected, as it is generally accepted that magnetic storms do occur almost simultaneously throughout the world [3].

Aurorae have been known to have zones of maximum frequency in the northern hemisphere for many years. Fritz [9] plotted curves of auroral zones in 1881. These have recently been revised, and similar curves plotted for the southern hemisphere by Vestine [19] in 1947.

Thus, for correlation studies of Aurora Borealis and Aurora Australis, one should compare observations for places in the same frequency belt. Since aurorae are usually, and most easily, observed in the evening hours, one should compare observations for places of approximately the same longitude. Auroral zones roughly follow the geomagnetic latitudes; thus, for any correlation studies, auroral observations from places of the same geomagnetic latitude and longitude should be compared.

Attempts have been made to compare occurrences of aurorae in the two hemispheres before, but with not much success. Some great displays have been recorded simultaneously in both northern and southern latitudes [5], [6], [8]. Loomis [15] found thirty-four between 1841 and 1848 for Hobart, Tasmania, and the United States. Davies [7] mentions a proposed comparison of times of auroral occurrence in the Antarctic in 1929 with those for northern stations. La Cour [14] suggested a world network of observing stations.

The literature surveyed afforded some 125 simultaneous (within 12 hours) occurrences of aurorae in similar middle north and south geomagnetic latitudes. However, the available published records were thought to be indeterminate respecting their completeness.

Due to the kind co-operation of Dr. I. L. Thomsen, Carter Observatory, New Zealand, and Dr. C. W. Gartlein, Cornell University, it has been possible to obtain complete auroral records from New Zealand from 1931 to the present date, and from the Lake Region of New York State from 1938 to 1946. It is chiefly these observations, then, for the years 1938 to 1946, inclusive, which have been used for these correlation studies. The use of these observations involves a compromise in geomagnetic longitude, but the geomagnetic latitudes (or relative frequency zones of aurorae) compare favourably. The frequency at Cornell is apparently greater because of the employment of a spectrograph which records an aurora at any time during the night [11]. The New Zealand observations are visual, but they are compiled from a large number of observers and hence are considered quite comparable in completeness with the Cornell observations, although the chance of observing an aurora in the early morning hours (local time) in New Zealand will be less than at Cornell. The following have also been available: (1) auroral observations from ionospheric observing stations in Canada supplied by Mr. F. T. Davies of the Defence Research Board of Canada; and (2) observations at Saskatoon submitted by Dr. B. W. Currie. The Canadian observations have not been used in this study, since some are for zones nearer the maximum auroral zone and others are somewhat less complete than the Cornell observations.

The times of occurrence of aurorae in the two "stations" permit only a few true coincidences. The time of usual occurrence in the north is between zero and nine hours (Greenwich Mean Time); the time of occurrence in the south is usually between seven and twelve hours (GMT). For this reason, for the present study, a coincidence is considered to have occurred if two aurorae (one Borealis, one Australis) occur on the same Greenwich date; for the most part, this means the two occur between zero and twelve hours (GMT).

Since the two "stations" are not of the same geomagnetic longitude, one might ask what is the time an aurora must endure in order that it be seen at different longitudes? However, if aurorae are caused by streams of radiation from the sun, these will sweep across the earth in some thirty-five seconds [3], so that the time taken for the earth to rotate need not be considered.

Results and Comments. All the Aurora Australis observations between 1938 and 1946, as supplied by Thomsen, have been compared with the observations of Aurora Borealis, as supplied by Gartlein.

For nine years (3,285 days), Aurora Borealis occurred on 1,316 individual days. This corresponds to a probability of there being an observation on any one day of 0.40. For the same period, Aurora Australis was observed on 863 individual days, or a probability of 0.26. Aurorae were observed on the same day as Borealis and Australis on 498 occasions, or a probability of an aurora being observed at both "stations" of 0.15.

Now one may consider the two phenomena as independent, and draw up a Euler-Venn table (see next page).

If the two phenomenon are independent, then the probabilities are as they occur in the table. On the basis of the results, it seems justifiable to say the two phenomena are not independent, for:

(1) 498 coincidences are noted, which is a probability of $\frac{498}{3,285}$ – 0.15; the Euler-Venn Table gives for this probability only 0.10;

(2) if the two phenomena are independent, then, of the aurorae observable as Aurora Australis, $\frac{.104}{.26}$ = 0.40 will be observable as Aurora Borealis; the comparison gives $\frac{498}{863}$ = 0.58 of Aurora Australis as Aurora Borealis.

Borealis Australis	p = 0.40	q = 0.60
p = 0.26	.104	.156
q = 0.74	.296	.444

It is concluded that there is a correlation between the Aurora Borealis and Aurora Australis.

It is noted that the number of possible coincidences is limited by the weather conditions, as well as by the possibility of an Aurora Australis being missed if it occurs in early morning (local time). Assessing the effect of weather conditions is difficult. Ithaca, New York, had the following number of cloudy days (days on which cloud cover was eight-tenths or more of the sky) [18]:

Year	*No. of cloudy days*
1938	213
1939	183
1940	203

This effect is somewhat overcome by the use of the patrol spectrograph which runs every night [11].

The following table compares intensities of the 498 coincident aurorae between 1938 and 1946. Aurorae Australis intensities are assigned 1, 2, 3, 4, in order of increasing intensity. Aurorae Borealis intensities are assigned 0, 1, . . . 9, in order of increasing intensity. Those in the queried row are unassigned. One might expect great intensity of one to correspond to great intensity of the other aurora. Such a correspondence is not strongly in evidence in the table.

Borealis / Australis	1, 2	3, 4	5, 6	7, 8, 9	Total
1	79	89	69	14	251
2	31	40	45	3	119
3	12	11	22	8	53
4	0	5	5	19	29
?	11	15	14	6	46
Total	133	160	155	50	498

Aurora Borealis observations (from Davies) for Ottawa, Ontario, for the years 1943 to 1949 have been studied. However, comparing these with the Cornell observations, it is found for 1945: (1) aurorae were observed on 65 days at Ottawa, on 145 days at Cornell; and (2) of the 34 coincidences found using the Cornell observations, 15 were found for Ottawa; i.e., 19 were missed. For these reasons, no attempt has been made to find coincidences between Aurorae Borealis obtained at Ottawa and Aurorae Australis observed in New Zealand in 1947, 1948, and 1949.

The authors wish to express their appreciation of the assistance and co-operation received from many workers in this field. They are especially indebted to Dr. I. L. Thomsen of Carter Observatory, New Zealand, and to Dr. C. W. Gartlein of Cornell University for their detailed records of auroral observations, and to Mr. Kenneth Parry for his assistance in this study.

This work was made possible by a grant in aid of research from the National Research Council of Canada.

REFERENCES

[1] H. Alfven. *On the theory of Aurora, the emission spectra of the night sky and Aurorae*, The Phys. Soc. (London, 1948).
[2] E. Bruche. Terr. Magn. and Atmos. Elec., vol. 36 (1931), 41.
[3] S. Chapman. *Theories of terrestrial and solar magnetism.* Dictionary of Applied Physics, vol. 2 (1922), 543.
[4] ——— *Theories of Aurorae, the emission spectra of the night sky and Aurorae*, The Phys. Soc. (London, 1948).
[5] ——— and J. Bartels. *Geomagnetism* (Oxford, 1940).
[6] B. W. Currie. J. Roy. Astron. Soc. Can., vol. 41 (1947), 249.
[7] F. T. Davies. Terr. Magn. and Atmos. Elec., vol. 36 (1931), 199.

[8] A. S. EVE. Nature, vol. 137 (1936), 813.
[9] H. FRITZ. *Das Polarlicht* (Leipzig, 1881).
[10] C. W. GARTLEIN. Sky and Telescope, vol. 1 (1941), 16.
[11] ——— Trans. Amer. Geoph. Union, 25th Annual Meeting (1944), 533.
[12] N. H. HECK. Sci. Amer., vol. 144 (1931), 153.
[13] E. W. HEWSON. Rev. Mod. Phys., vol. 9 (1937), 403.
[14] D. LACOUR. Int'l Geod. and Geoph. Union, section of Terr. Magn. and Elec., Bull. 8 (Stockholm, 1931).
[15] E. LOOMIS. Amer. J. Sci. (3), vol. 5 (1873), 245
[16] M. NICOLET. *Deductions regarding the state of the high atmosphere, the emission spectra of the night sky and Aurorae*, The Phys. Soc. (London), 1948.
[17] W. J. ROONEY. Terr. Magn. and Atmos. Elec., vol. 39 (1934), 103.
[18] U. S. Meteorological Yearbook (1938, 1939, 1940); U.S. Dept. of Agric. (1938, 1939); U. S. Dept. of Commerce (1940); Washington.
[19] E. H. VESTINE. *The geomagnetic field* . . ., Carnegie Inst. of Washington, Dept. of Terr. Mag. (1947), 385.

TRANSACTIONS OF

THE ROYAL SOCIETY OF CANADA

SECTION III

CHEMICAL, MATHEMATICAL,
AND
PHYSICAL SCIENCES

THIRD SERIES—VOLUME XLV—SECTION III

JUNE, 1951

56673

Ottawa
The Royal Society of Canada
1951

TABLE OF CONTENTS

Fission Product Yields and Shell Structure in Atomic Nuclei. Presidential Address. By H. G. THODE, F.R.S.C. 1

Riesz Methods of Summation and Orthogonal Series. By G. G. LORENTZ 19

Means of Fourier Constants. By G. M. PETERSEN 33

The Dielectric Constants of Binary Mixtures. By J. W. TOMECKO and W. H. HATCHER, F.R.S.C. 39

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLV : SERIES III : JUNE, 1951
SECTION III

PRESIDENTIAL ADDRESS

Fission Product Yields and Shell Structure in Atomic Nuclei

H. G. THODE, F.R.S.C.

DURING the past six years we at McMaster have been interested in nuclear chemical problems, particularly those relating to isotope identification and abundance measurements [1], [2], [3], [4], [5], [6]. In this address, I wish to review this work briefly and to discuss some of the more recent developments.

In the fission of U^{235} with thermal neutrons, many isotopes of the fission elements are produced. The fission process can be represented as follows:

$$
\begin{aligned}
U^{235} + n \rightarrow U^{236*} \rightarrow {}_{Z_1}A^{m_1} &+ {}_{Z_2}B^{m_2} + vn \\
Z_1 + Z_2 &= 92 \\
v + m_1 + m_2 &= 236
\end{aligned}
$$

where Z is the atomic number, m the mass, and v the number of neutrons emitted.

The primary products A and B have more neutrons for their charge than stable nuclei and undergo β disintegration in several steps, ending in stable nuclear structure or stable isotopes (as found in nature). Some sixty-four such fission chains are known [7].

It was clear from the first fission product studies that the mass spectrometer would eventually be useful in separation problems, mass identifications, and isotope abundance measurements. Actually, the mass spectrometer provides the only direct means of identifying and measuring stable isotopes, which are end products of fission product chains. However, the early work on fission products involved such small quantities of material that only radiochemical methods were considered sensitive enough to measure and follow the radioactive isotopes.

Early in this work Drs. Arrol, Chackett, and Epstein succeeded in measuring small volumes of nobel gases extracted from uranium irradiated with neutrons [8]. At that time, mass spectrometer studies on such small samples were considered out of the question. However, a discarded sample containing about 0.001 cubic mm. of mixed rare gases was found to be sufficient. With this sample we were able to show that the sample contained A and Kr in addition to xenon, to identify the fission product isotopes of Xe and Kr and to measure their abundances. Fig. 1 shows a mass spectrogram of normal xenon, which is a mixture of nine isotopes. Fig. 2 shows that in fission product xenon there are only four isotopes, namely, Xe^{131}, Xe^{132}, Xe^{134}, and Xe^{136}. These fission isotopes are stable end products of fission product chains. The mass spectrometer studies made it possible to relate these isotopes to the radioactive members of the chains and thereby assign masses to the respective chains with certainty. Further, since the chains were allowed to decay to stable xenon before extraction, the relative abundances of the isotopes gave direct information about the cumulative fission yields of the different mass chains.

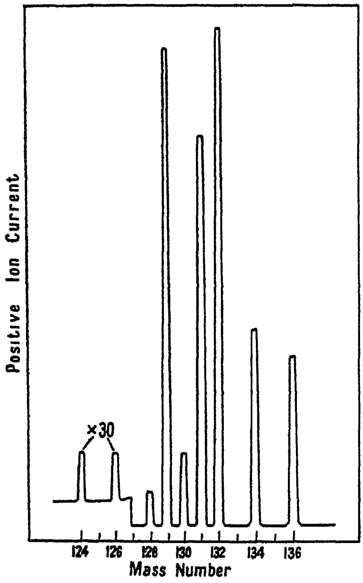


FIGURE 1.—Isotopes of normal xenon.

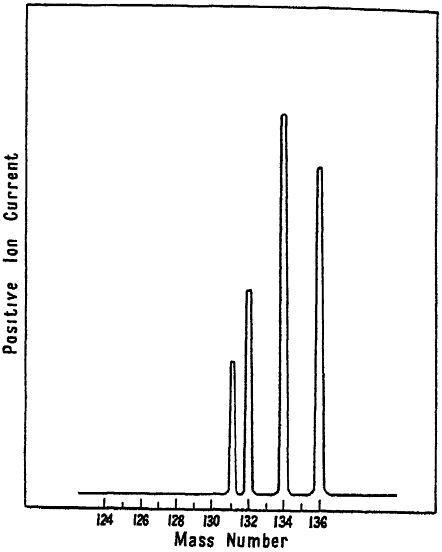


FIGURE 2.—Isotopes of fission product xenon.

Fig. 3 shows the mass pattern obtained for normal Kr, which is a mixture of six isotopes, and Fig. 4, the mass pattern for fission product krypton. Here again, these isotopes could be related to known radioactive isotopes of Br, which decay to stable Kr, and their masses assigned. The Kr^{85} isotope, which does not appear in natural-occurring Kr, is radioactive. It was found to be present in all fission product samples, but its relative concentration was less in older samples. Its half-life of 9.4 years was determined [1] by measuring its concentration relative to a stable isotope of two different times. Actually, the existence of this long-lived isotope of Kr was not suspected because a four-hour Kr^{85} had already been shown to exist, and this isotope would have completely decayed in samples six months to a year old. It turns out that the two are independent isomers of Kr^{85} [1], [7].

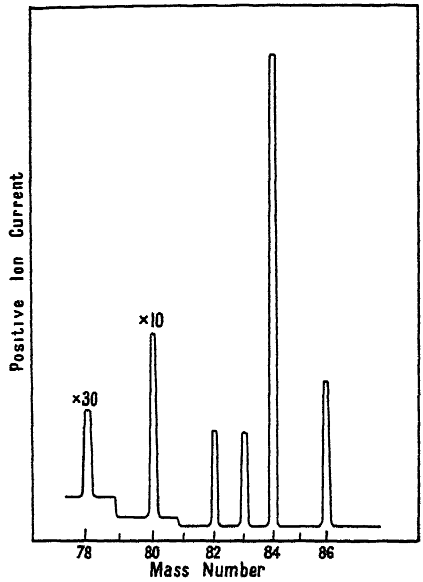


FIGURE 3.—Isotopes of normal krypton.

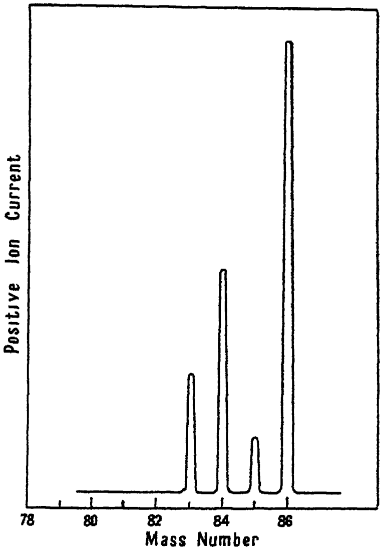


FIGURE 4.—Isotopes of fission product krypton.

Fig. 5 shows the fission chains leading to isotopes of Xe and some of the information we now have about these chains. The absence of Xe^{129} in fission product material is of considerable interest since Te^{129}, which is β active, has been positively identified. This means that undoubtedly I^{129} is formed, although so far it has escaped detection. Absence of Xe^{129} would certainly indicate that I^{129} has a long half-life and blocks the chain at this point. I will mention this chain later in connection with spontaneous fission.

In the past year, we have extended this work to include the cesium and the rubidium isotopes [9]. It can be seen from Fig. 5 that three chains lead to stable or long-lived isotopes of cesium. Up until that time, all work in our laboratory had involved gaseous samples. It was felt that much larger samples would be required to produce ions in a mass spectrometer where solid samples were involved. Actually, this is not the case for the alkali metals. It was found that samples of the order of 1 microgram of Cs and 0.01 to 0.1 micrograms of Rb were quite adequate for isotope identification and measurement. This means that our mass spectrometer studies of fission products can be extended to many elements without having to use remote control methods to handle large quantities of irradiated uranium.

The isotopes of Cs and Rb are of particular interest since their measurement gives information about fission mass chains adjacent to those of Xe and Kr in the neighbourhood of the 50 and 82 neutron

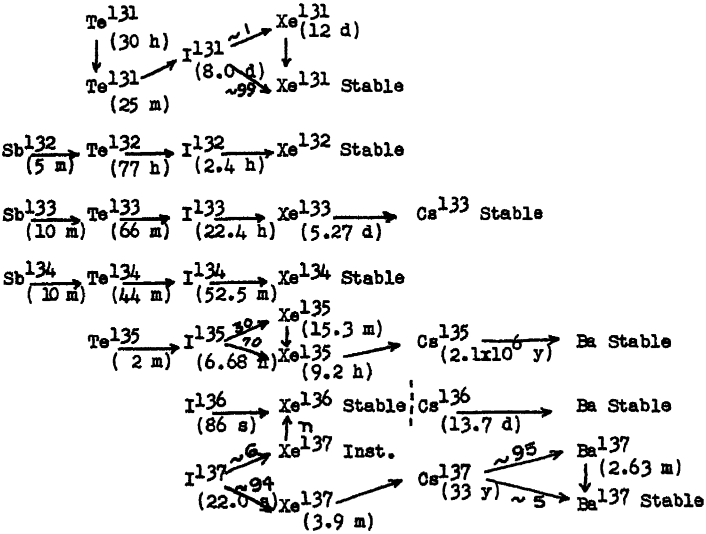


FIGURE 5.—Fission chains leading to xenon, cesium, and barium.

shells where perturbations occur in the fission yields. It should be pointed out here that Inghram and Hess and Reynolds [10] had investigated the Cs isotope previous to us. However, they had been unable to check their results for normal Cs^{133} contamination from the reagents and uranium, and their high result for Cs^{135} could not be explained. Our recent work on Cs was, therefore, necessary.

Fig. 6 shows the fission chains leading to stable or long-lived isotopes of Kr. It is seen from the branching that the two isomers of Kr^{85} decay independently to the next element Rb.

Perhaps the most important aspect of fission product studies at the present time involves yields, or the percentage of fission which leads to a certain pair of fission masses. Both primary fission yields and cumulative yields for each mass are important.

The great mass of yield data collected from the various workers in the field is shown in Fig. 7 with the accumulative yield for each mass plotted against mass [7]. This is the familiar double-hump fission yield curve, showing the distribution of fission fragments according to mass. This asymmetrical fission of uranium could not have been predicted with our present information about atomic nuclei. All theoretical calculations carried out for the fission process predict that symmetric, equal mass fragments should be produced. The asymmetric fission is undoubtedly due to the existence of shell structure in atomic nuclei, about which much has been written in the past few years [11], [12], [13], [14]. The two broad peaks in the yield curve fall in the region of nuclei characterized by 50 and 82 neutrons, which seem to be particularly stable configurations. The shell structure would appear to distort the dependence of binding energy on the number of particles so that more energy is released by dividing assymmetrically.

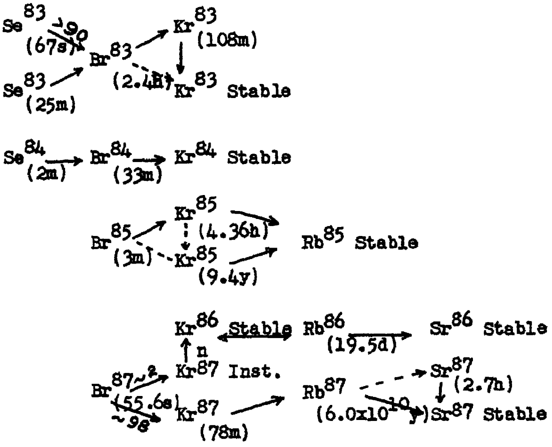


FIGURE 6.—Fission chains leading to krypton and rubidium.

Although there is no adequate theory for shell structure, the periodicity of properties of nucleides as a function of neutron and proton numbers and the abundances of isotopes prove beyond any doubt that we encounter saturated nucleonic shells similar to electronic shells, which are responsible for the periodic properties of atoms. Fig. 8 illustrates the phenomenon for atomic nuclei [11]. The relative

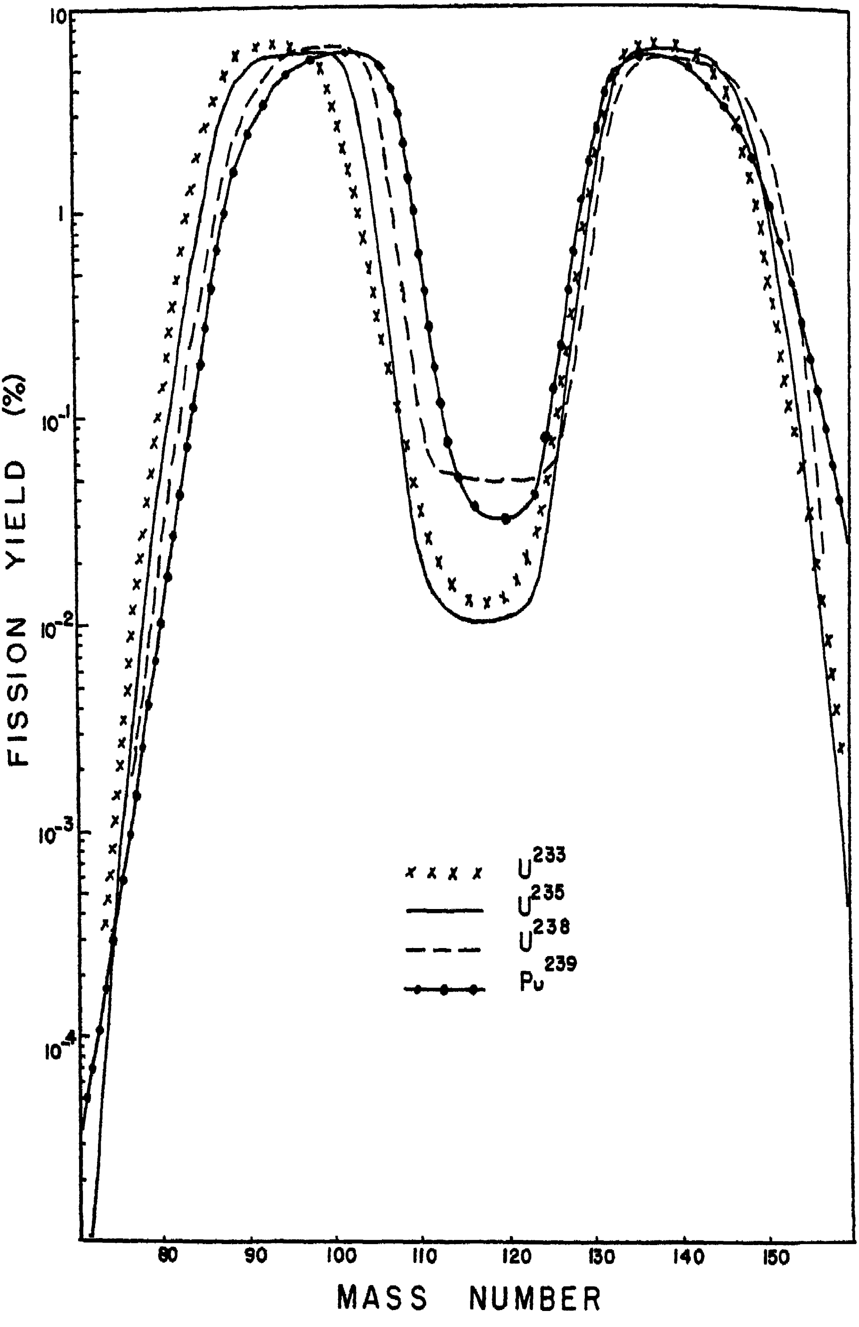


FIGURE 7.—Mass yield curves for four fissionable nucleides.

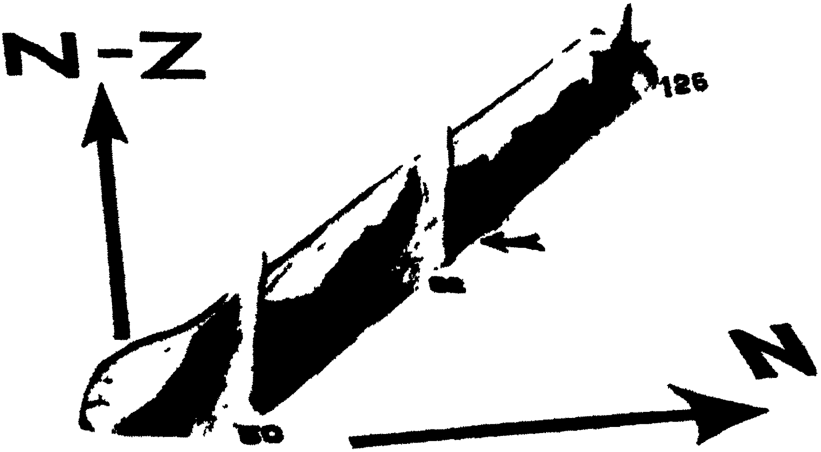


FIGURE 8.—A plot of the percentage abundance of isotopes in [N, N—Z] plane showing the distinct maxima at N = 50 and N = 82, and the rise towards the maximum at N = 126.

percentage abundances of the isotopes of the even elements are plotted as a mountain range in the plane (N-Z, N) from N = 30 to N = 128. The peculiar nature of the neutron numbers 50 and 82 is quite evident from the plot. It will be recalled that the products of fission cluster about these neutron numbers rather than about a mid point.

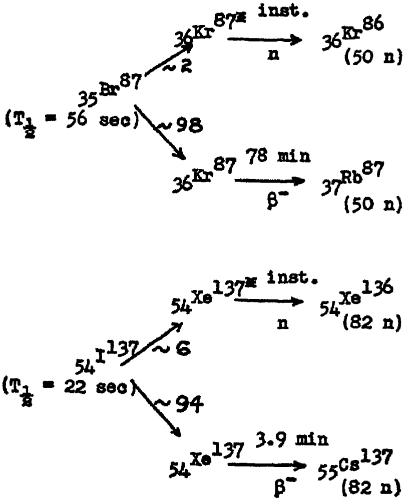


FIGURE 9.—Delayed neutron activity.

Fig. 9 shows another type of spontaneous disintegration where a single nucleon separates from the main body of nucleons—in other words, instantaneous neutron emission of some of the fission fragments [7]. The two isotopes which emit delayed neutrons that are so important to controlled chain reactions are Kr^{87} and Xe^{137}, which have 51 and 83 neutrons respectively. The neutron activity is delayed, depending on the decay of the Br^{87} and I^{137}. The 51st and 83rd neutrons must, therefore, be particularly weakly bound; otherwise, neutron emission could not take place. If 50 and 82 neutrons form closed shells, then one would expect the binding energy of the 51st and 83rd neutrons to be somewhat low.

Most fission yields have been determined by radiochemical methods and are probably only good to ±10 per cent. However, since mass spectrometer abundance data are good to ±0.1 per cent, relative fission yields can be obtained, using mass spectrometer methods, with high precision and accuracy. Fig. 10 gives a summary of mass spectrometer fission yield data for the Xe isotopes obtained to date. It is seen that the yields for masses 133 and 134 are abnormally high and do not fall on the smooth mass fission yield curve. This was the first evidence for "fine structure" in the mass yield curve. Glendenin [16] has suggested that these abnormal yields may be due to closed shell effects. He postulates that a primary fission product (which has already emitted the usual number of prompt neutrons) containing one neutron in excess of the closed shells (51 and 83) will often emit this

TABLE I [4]

$$Sn^{131} \rightarrow Sb^{131} \rightarrow Te^{131} \rightarrow I^{131} \rightarrow \frac{Xe^{131}}{\text{Stable}}$$

$$Sn^{132} \rightarrow Sb^{132} \rightarrow Te^{132} \rightarrow I^{132} \rightarrow \frac{Xe^{132}}{\text{Stable}}$$

$$\overset{1\uparrow n}{Sn^{133}} \rightarrow Sb^{133} \rightarrow Te^{133} \rightarrow I^{133} \rightarrow Xe^{133} \rightarrow \frac{Cs^{133}}{\text{Stable}}$$

$$\overset{2\uparrow n}{Sb^{134}} \rightarrow Te^{134} \rightarrow I^{134} \rightarrow \frac{\text{Stable}}{Xe^{134}}$$

$$Sb^{135} \rightarrow \overset{3\uparrow n}{Te^{135}} \rightarrow I^{135} \rightarrow Xe^{135} \rightarrow Cs^{135} \rightarrow \frac{Ba^{135}}{\text{Stable}}$$

$$Sb^{136} \rightarrow Te^{136} \rightarrow \overset{4\uparrow n}{I^{136}} \rightarrow \frac{Xe^{136}}{\text{Stable}}$$

$$Te^{137} \rightarrow I^{137} \rightarrow \overset{5\uparrow n}{Xe^{137}} \rightarrow Cs^{137} \rightarrow \frac{Ba^{137}}{\text{Stable}}$$

$$Te^{138} \rightarrow I^{138} \rightarrow Xe^{138} \rightarrow \overset{6\uparrow n}{Cs^{138}} \rightarrow \frac{Ba^{138}}{\text{Stable}}$$

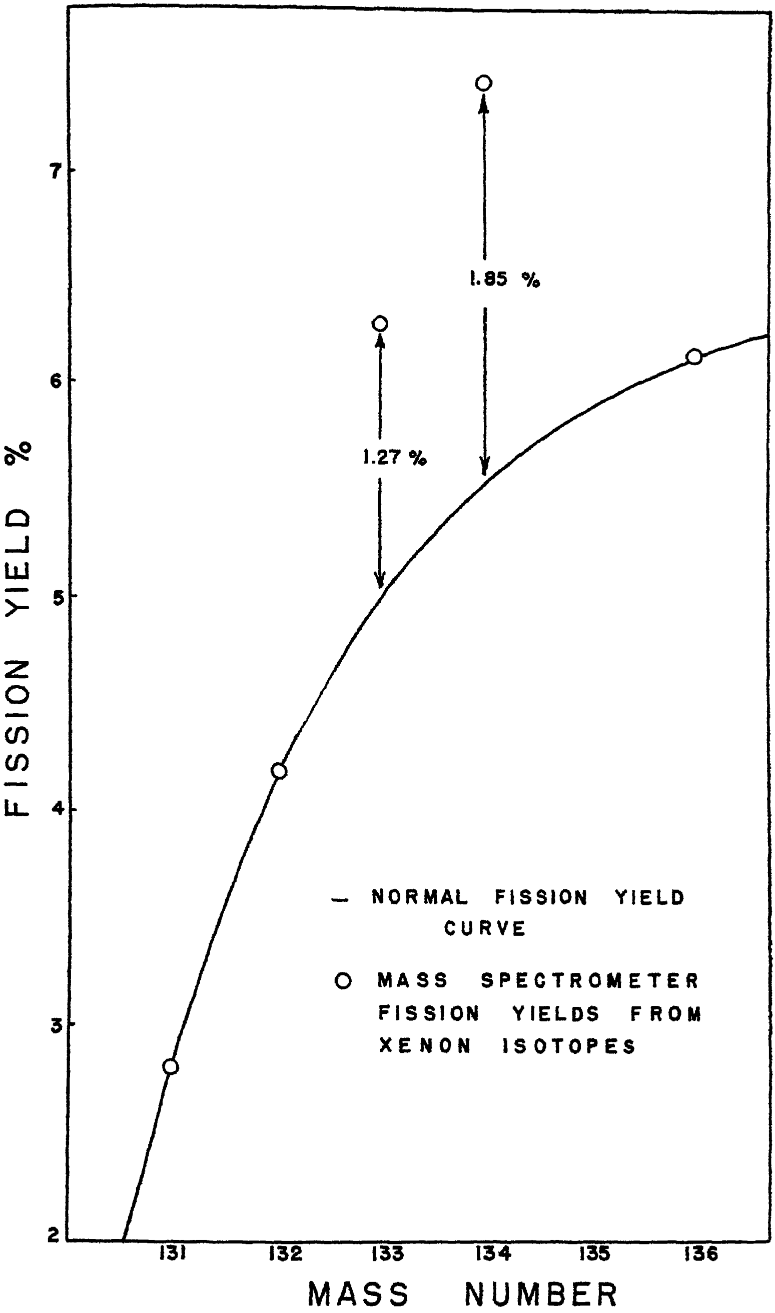


FIGURE 10.—Mass fission yield curve, showing fine structure.

more loosely bound neutron immediately rather than emit β-particles or γ -rays as in the ordinary case. This process could then account for abnormal fission yields in the neighbourhood of nucleides with 50 and 82 neutrons. Table I illustrates the chain branching in the region of 82 neutrons as proposed by Glendenin. This scheme would explain qualitatively the abnormal yields of Xe^{131}, I^{136}, and the abnormal yield of Xe^{133} reported in this paper. For example, the yield of the 134 chain is increased by the reaction

$$Te^{135}(83 \text{ neutrons}) \rightarrow Te^{134}(82 \text{ neutrons}) + n \qquad (3),$$

and is decreased by the reaction

$$Sb^{134}(83 \text{ neutrons}) \rightarrow Sb^{133}(82 \text{ neutrons}) + n \qquad (2).$$

If Te^{135} has a higher primary yield than Sb^{134}, then the 134 chain will gain more than it loses by this mechanism and its yield will be high [4].

Table I shows the side reactions of this type numbered from 1 to 6. If it is assumed that the fine structure in the mass yield curve is due to these side reactions, then the extent of some of these reactions can be established from the experimental results so far obtained. For example, the mass spectrometric yield data for Xe^{131} and Xe^{132} are in accord with the mass yield curve; the primary yield of Sn^{133} must

TABLE II

FISSION YIELDS

(U^{235} + n)

Mass chain	*Mass yield curve value*	*Mass spectrometer fission yield*
83	0.53	0.59
84	0.90	1.09
*85	1 30	.32
86	2.0	2.09
131	2.8	2.8
132	4.2	4.2
133	5.0	6.3
134	5.5	7.4
135	5.9	6.0
136	6.1	6.05
137	6.2	5.9
		(Normalized to 2.8 per cent at 131)

*Only 9 4 years $K^{85}r$ branch of chain therefore be small and the extent of reaction (1) approximately zero. Reaction (2), on the other hand, must amount to 1.3 per cent of total fission since the Xe^{133} yield is found to be high by this amount. In the same way, the extent of reaction (3) can be established [4].

Our most recent yield data are given in Table II. Unfortunately, the values at the present time are relative and the absolute values depend on taking the radiochemical yield of the mass 131 chain as a standard. What is needed, however, is a re-determination of the absolute yield of Xe by some entirely new method. The mass spectrometer might very well be used. The method would involve irradiating uranium in a nuclear reactor at a point where the neutron flux could be determined, extraction and measurement of volume of Xe produced in fission, and finally, an isotopic abundance measurement on the Xe sample.

SPONTANEOUS FISSION

Petrzhak and Flerov [17] first reported evidence of the spontaneous fission of uranium in 1941. Since then, spontaneous fission has been studied by a number of workers in the field, and half-lives from 1.9×10^{15} to 1.3×10^{16} years have been reported.

Recently, we have extracted samples of Xe and Kr gas from pitchblende in amounts up to 0.0001 cubic cm. of gas at N.T.P. [5]. These have been investigated with a mass spectrometer, and the results are given in Figs. 11 and 12. It is clear that these mass patterns are not those of natural-occurring Xe and Kr. However, there are some differences between spontaneous fission and the fission isotope pattern for $U^{235} + n$. First of all, Xe^{129} appears. This must be formed by the decay of I^{129}, which we predicted had a very long half-life. However, the pitchblende sample was 10^9 years old and some decay might be expected. Secondly, the fine structure is different. Xe^{134} is no longer higher in concentration than Xe^{136}. The Kr mass pattern is similar to that for $U^{235} + n$ fission, except that Kr^{85} does not appear. This is to be expected since it has a 9.4 year half-life, which is short compared to 10^9 years—the age of the sample. Fig. 13 shows a plot of the mass spectrometer yield data normalizing to 6.1 per cent at Xe^{136}.

Here again, we have asymmetric fission with peaks about the 50 and 82 neutron shell structures, although the peaks are somewhat narrower. The question arises as to whether those products are due to spontaneous fission or to *n* fission of U^{235}. The latest values for the amount of plutonium in pitchblende [18] make it possible to calculate the average neutron flux over the past few hundred thousand years,

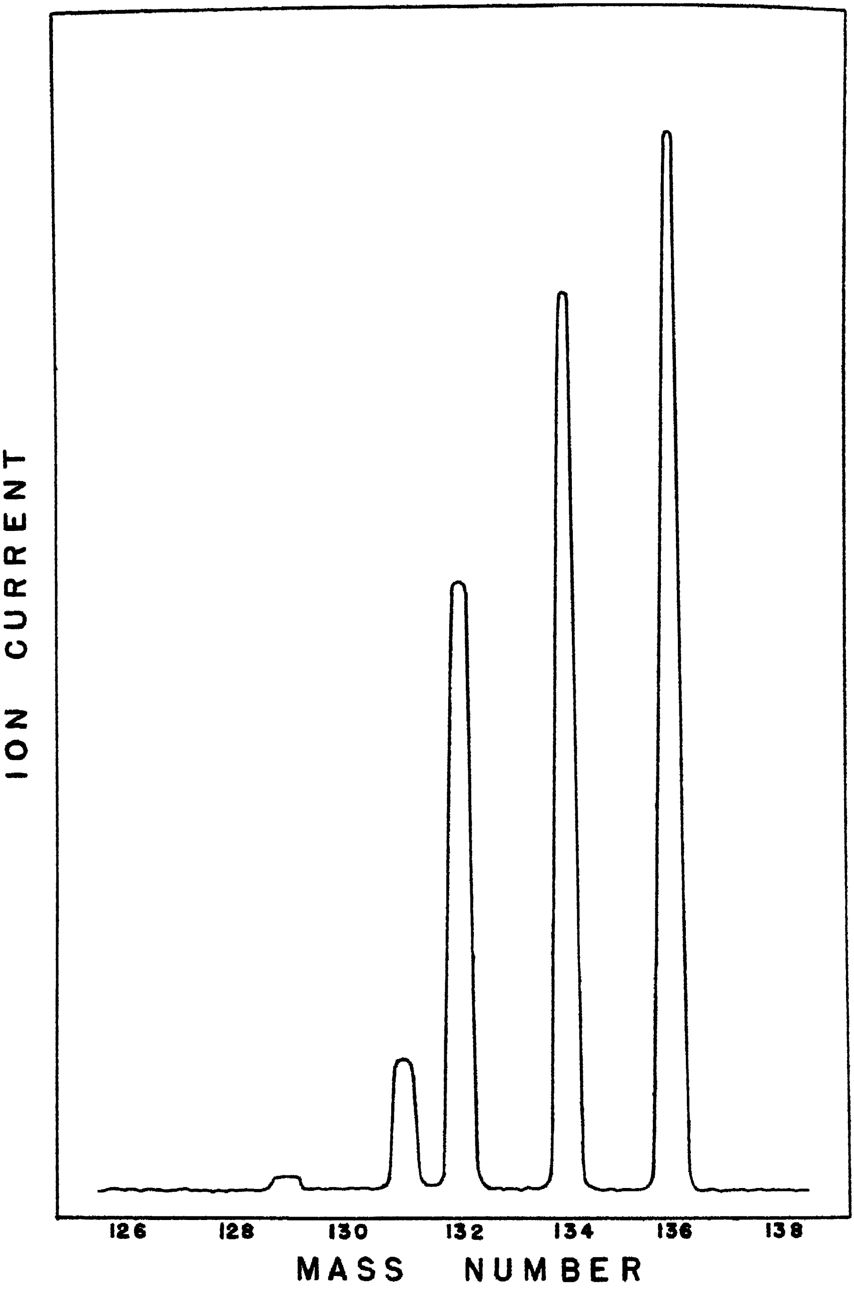


FIGURE 11 —Xe isotopes from spontaneous fission.

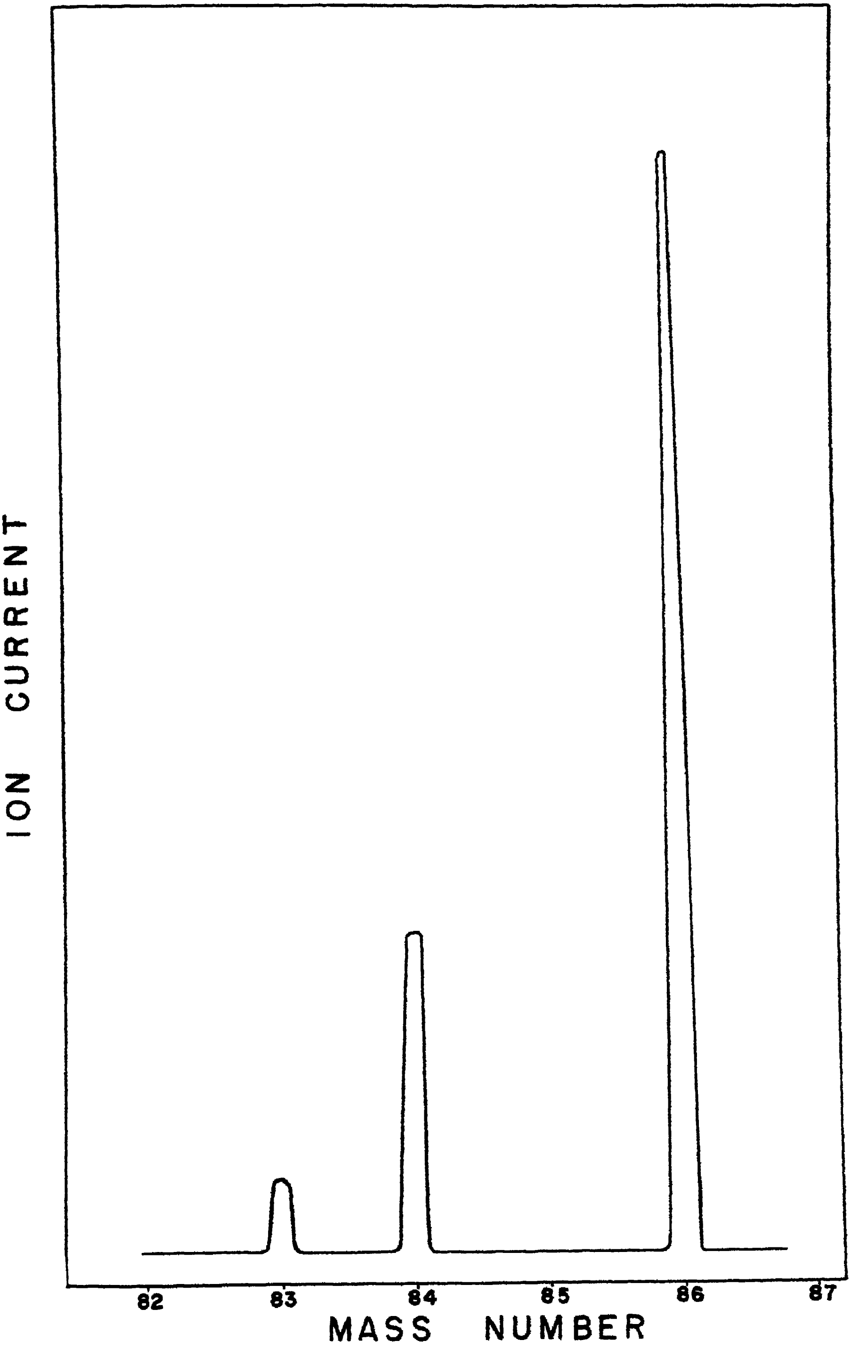


FIGURE 12.—Kr isotopes from spontaneous fission.

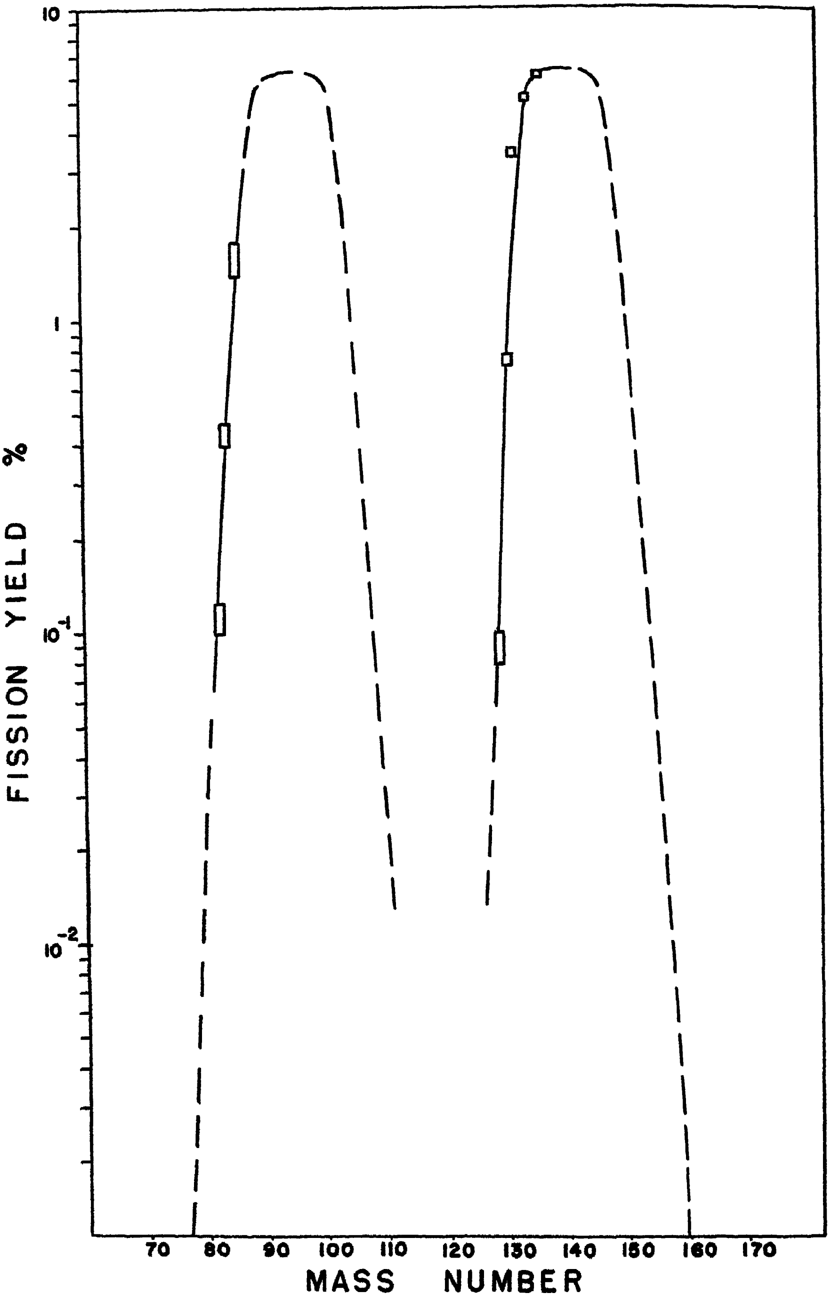


FIGURE 13.—Fission yield curve, spontaneous fission of uranium.

assuming an equilibrium value of plutonium in the pitchblende. This turns out to be a thousandfold too low to account for the amount of fission which has taken place. Further, the amount of the fission products found agrees roughly with the half-life reported for spontaneous fission.

According to the latest information, the spontaneous fission is probably due to both U^{235} and U^{238} [19]—at the present time, ⅓ to ¼ U^{235} and ⅔ to ¾ U^{238}.* However, 10^9 years ago the pitchblende contained more U^{235} and the proportion of spontaneous fission due to U^{235} would be greater. The older sample would, therefore, have a greater proportion of U^{235} fission products.

Fission induced by high energy neutrons tends to give symmetrical fission where the excess energy is large as compared to binding energies due to closed shells. With low energy neutrons, we get the characteristic asymmetrical fission. Finally, with spontaneous fission, the average energy available for fission will be less and the distribution of fission products about the 50 and 82 neutron number should be even narrower. This would seem to be borne out by the results obtained to date.

Fig. 7 shows the shift in the mass yield curve as the mass of the fission element is changed. It would appear that the shift in yield values is largely in the light mass group. This would mean a large change in the Xe/Kr ratio. For U^{235} this ratio is 5 to 1. For U^{238} this ratio should be even larger, as the Kr concentrations should be much lower. Finally, for U^{233} the ratio should be about 2 to 1. Thus it should be possible to establish the mass of the element undergoing fission from a determination of the ratio of Xe to Kr in the fission products.

Finally, Fig. 14 shows the fission yields obtained for two samples of pitchblende of different ages. The ages of the pitchblende samples were determined in two ways: first, by measuring the ratio of the lead isotopes 206 and 207; and secondly, the ratio of lead to uranium. According to the latest results from Oak Ridge, I^{129} has a half-life of 1.72×10^7 years [20]. This means that most of the I^{129} will have decayed to Xe^{129} for these very old samples.

It is seen that for the younger sample of pitchblende, the yields are increasingly higher as we proceed from masses 134 to 129. The reason for this discrepancy between the mass patterns for the fission products for the two pitchblende samples is not known. However, it

---

*Information which has come to hand since this address was given indicates that the U^{235} contribution to spontaneous fission must be negligible even in the older samples. Emilio Segrè, University of California, private communication.

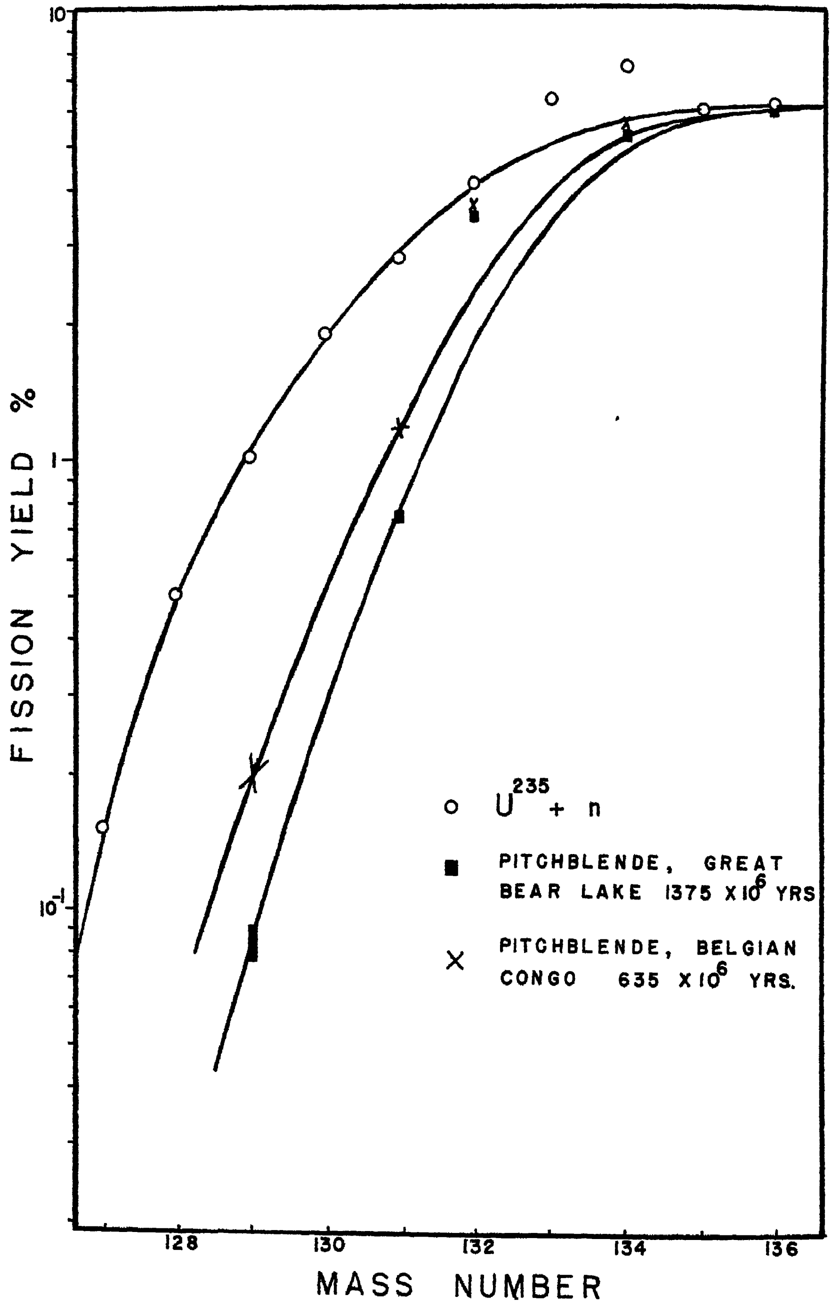


FIGURE 14.—Comparison of fission yield curves for U^{235} + n and spontaneous fission.

must be connected in some way with the different ages of the two samples. If we have a mixture of U^{238} and U^{235} fission products, then the older samples will be richer in fission products from U^{235}. If further studies bear this out, then a new geological time clock might be worked out.

It is clear that the mass spectrometer provides a powerful tool for nuclear studies and will be used more and more in the future.

REFERENCES

[1] H. G. THODE and R. L. GRAHAM. Can. J. Res., vol. A25 (1947), 1.

[2] H. G. THODE. Nucleonics, vol. 3 (1948), 14.

[3] J. MACNAMARA, C. B. COLLINS, and H. G. THODE. Phys. Rev., vol. 75 (1949), 582.

[4] ———Phys. Rev., vol. 78 (1950), 129.

[5] J. MACNAMARA and H. G. THODE. Phys. Rev., vol. 80 (1950), 471

[6] ———Phys. Rev., vol. 80 (1950), 296.

[7] Plutonium Project Report, J. Amer. Chem. Soc., vol. 68 (1946), 2411.

[8] W. J. ARROL, K. F. CHACKETT, and S. EPSTEIN. Can. J. Res., vol. B27 (1949), 757.

[9] D. WILES. M.Sc. thesis, McMaster University (1950); B. W. SMITH. M.Sc. thesis, McMaster University (1951).

[10] M G. INGHRAM, D. C. HESS, and J. H. REYNOLDS. Phys. Rev., vol. 76 (1949), 1717.

[11] J. H. BARTLETT. Nature (1932), 130, 165; also W. D. HARKINS. J. Franklin Institute (1923), 194, 645.

[12] M. G. MAYER. Phys. Rev., vol. 74 (1948), 235.

[13] E. FEENBERG and K. C. HAMMACH. Phys. Rev., vol. 75 (1949), 1877.

[14] L. W. NORDHEIM. Phys. Rev., vol. 75 (1949), 1894.

[15] G. GAMOW and C. L. CRITCHFIELD. *Theory of atomic nucleus*, Oxford University Press (1949).

[16] L. E. GLENDENIN. Phys. Rev., vol. 75 (1949), 337.

[17] K. A. PETRZHAK and G. N. FLEROV. J. Phys. U.S.S.R., vol. 3 (1940), 275.

[18] PEPPARD, STUDIER, GERGEL, MASON, SULLIVAN, and MECH. Paper presented at 118th meeting of American Chemical Society, Chicago (Sept., 1950).

[19] R. E. LAPP and H. L. ANDREWS. *Nuclear radiation physics*, New York (1948), 357.

[20] S. KATCOFF, O. A. SCHAEFFER, and J. M. HASTINGS. Phys. Rev., vol. 82 (1951), 688.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLV : SERIES III : JUNE, 1951
SECTION THREE

Riesz Methods of Summation and Orthogonal Series*

G. G. LORENTZ

Presented by R. L. JEFFERY, F.R.S.C.

IN this paper we wish to show the usefulness of Riesz means for the study of orthogonal series and of general methods of summation. Our results show again that applications of Riesz methods $R(\lambda_n, k)$ are by no means limited to Dirichlet series for which they were invented. In the theory of Fourier series their importance is partly veiled by the equivalence of means $R(n, k)$ to the considerably simpler Cesàro means. For our problems, however, the case $\lambda_n = n$ is of no particular importance.

By an orthogonal series we understand

$$(1) \qquad \sum_{n=0}^{\infty} a_n \phi_n(x), \quad \sum a_n^2 < +\infty,$$

where $\phi_n(x)$, $n = 0, 1, 2, \ldots$ are orthogonal and normalized on an interval (a, b). Section 1 is devoted to theorems on the Riesz summability of series (1), which generalize the known results about the C_k-summability (see, for instance, Kaczmarcz and Steinhaus [4]). We obtain some applications in §2, where we show, for instance, that for any orthogonal system $\phi_n(x)$ there is a method $R(\lambda_n, 1)$ effective with respect to all orthogonal series of the $\phi_n(x)$, just as C_1 is Fourier effective in L^2. In §3 we prove theorems on inclusion of the method $R(\lambda_n, 1)$ in a general method A. A relation to the "summability functions," discussed in [7] and [8], is found. Combining the results of §1 and §3 we obtain in §4 partial generalizations of theorems of §1 to general methods of summation. Finally, we also give simple proofs of some known theorems, for instance, of a theorem of Menchoff in §4. We are concerned mainly with Riesz means $R(\lambda_n, k)$ of order $k = 1$, but in many cases a generalization to $k > 0$ is possible.

1. Riesz summability of orthogonal series and convergence of subsequences of partial sums. Let λ_n be a sequence of positive numbers increasing to $+\infty$, $k > 0$, and $\lambda(\omega)$ a continuous and monotone extension of λ_n for all $\omega \geqslant 0$. We denote by $\Lambda(\omega)$ the function inverse to $\lambda(\omega)$. A series $\sum u_n$ is called $R(\lambda_n, k)$-summable to s if

---

*This investigation was carried out while the author held a fellowship at the Summer Institute of the Canadian Mathematical Congress in 1950.

$$\sigma_\omega = \sigma_\omega^k = \lambda(\omega)^{-k} \sum_{\nu \leqslant \omega} [\lambda(\omega) - \lambda_\nu]^k u_\nu \to s \tag{2}$$

for $\omega \to \infty$. We shall be interested mainly in the case $k = 1$. It is well known that in this case $R(\lambda_n, 1)$ is equivalent to the "method of arithmetic means" defined by

$$\sigma_n = \lambda_n^{-1} \sum_{\nu \leqslant n} (\lambda_n - \lambda_\nu) u_\nu \to s. \tag{3}$$

Therefore, the partial sums of (3) will be also called $R(\lambda_n, 1)$-sums.

In this section we shall often be concerned with convergence a.e. (almost everywhere) of a sequence $s_{n_\nu}(x)$ of the partial sums

$$s_\omega(x) = \sum_{n \leqslant \omega} a_n \phi_n(x),$$

where the n_ν need not be integers. There are certain special sequences n_ν associated with a given sequence λ_n. The following conditions on n_ν will be used:

$$\omega(n) \leqslant \log^C \lambda_n \qquad (C > 0), \tag{A}$$

where $\omega(n)$ is the counting function of the n_ν, that is, the number of n_ν for which $n_\nu \leqslant n$.

$$\lambda(n_{\nu+1})/\lambda(n_\nu) \geqslant q > 1. \tag{B}$$

$$n_\nu \text{ is the sequence } m_\nu = \Lambda(2^\nu). \tag{C}$$

Clearly (C) implies (B). Again, (B) implies (A), for it follows from (B) that $\lambda(n_\nu) \geqslant C_1 q^\nu$ and therefore $\omega(n)$ will be

$$\omega(n) \leqslant \frac{\log \lambda(n) - \log C_1}{\log q} \leqslant \log^C \lambda(n).$$

Our first theorem contains the theorem of Menchoff-Rademacher:

Theorem 1. *$s_{n_\nu}(x)$ converges a.e. if n_ν satisfies* (A) *and if*

$$\sum a_n^2 (\log \log \lambda_n)^2 < +\infty. \tag{5}$$

Proof. Putting

$$r_m = \sum_{n>m} a_n^2,$$

we have

$$\sum^{\infty} \frac{1}{\nu} r_{n_\nu} \log \nu = \sum^{\infty} a_n^2 \sum \frac{\log \nu}{\nu} \leqslant C_1 \sum^{\infty} a_n^2 \log^2 \omega(n)$$

$$\leqslant C_2 \sum_{n=1}^{\infty} a_n^2 (\log \log \lambda_n)^2 < +\infty.$$

It is now sufficient to apply a well-known lemma [4, p.164].

THEOREM 2. *If* $\{n_\nu\}$ *has the property* (B), *then*

$$\sigma_{n_\nu}(x) - s_{n_\nu}(x) \to 0 \qquad a.e. \tag{6}$$

Proof. We have

$$\sigma_\omega(x) - s_\omega(x) = -\frac{1}{\lambda(\omega)}\sum_{n\leqslant\omega}\lambda_n a_n\phi_n(x),$$

and putting

$$\delta_\omega = \int_a^b (\sigma_\omega - s_\omega)^2 dx,$$

$$\sum \delta_{n_\nu} = \sum_{\nu=1}^{\infty}\frac{1}{\lambda(n_\nu)^2}\sum_{n\leqslant n_\nu}\lambda_n^2 a_n^2 = \sum_n a_n^2\lambda_n^2\sum_{n_\nu\geqslant n}\frac{1}{\lambda(n_\nu)^2}$$
$$\leqslant C_3\sum a_n^2 < +\infty.$$

This proves our result, which is a generalization of a theorem of Kolmogoroff [4, p.161].

THEOREM 3. *An orthogonal series* (1) *is a.e.* $R(\lambda_n, 1)$*-summable if and only if the sequence of partial sums* $s_{n_\nu}(x)$ *is a.e. convergent, where* n_ν *satisfies* (B) *or* (C).

For $\lambda_n = n$, when the $R(\lambda_n, 1)$ method reduces to C_1 this theorem is due to Kaczmarz. We omit the proof, since it follows the known lines (see [4, p.190]), remarking only that the lemma to be used in it takes the form:

$$\sum_{\nu=1}^{\infty}\nu[\sigma_{\Lambda(\nu+1)}(x) - \sigma_{\Lambda(\nu)}(x)]^2 \qquad \text{a.e.}$$

We now obtain, combining Theorems 1 and 3:

THEOREM 4. *If*

$$\sum a_n^2(\log\log\lambda_n)^2 < +\infty, \tag{7}$$

then $\sum a_n\phi_n(x)$ *is a.e.* $R(\lambda_n, 1)$*-summable.*

Taking $\lambda_n = e^n$, we obtain the theorem of Menchoff-Rademacher; and for $\lambda_n = n$, the theorem of Menchoff-Kaczmarz that

$$\sum a_n^2(\log\log n)^2 < +\infty$$

implies that $\sum a_n\phi_n(x)$ is a.e. C_1-summable. One finds easily, by reduction to the case

$$\sum a_n'^2\log^2 n < +\infty,$$

that (7) is the best possible condition of this kind.

To prove a theorem on strong $R(\lambda_n, 1)$ summability, we need the following lemma, for $k = 1$. In view of later applications, we prove it here for any $0 < k \leqslant 1$.

LEMMA 1. *If* $\Delta\lambda_n$, $\lambda_n/\Delta\lambda_n$ *are increasing and* $0 < k \leqslant 1$, *if further* $\sum c_n$ *converges, then the sequence* $(\lambda_n/\Delta\lambda_n)^k c_n$ *is* $R(\lambda_n, k)$*-summable to* 0.

Proof. Let $\mu_n = (\lambda_n/\Delta\lambda_n)^k$, then the $R(\lambda_n, k)$-mean σ_ω of the sequence $\mu_n c_n$ takes the form

$$\sigma_\omega = \frac{1}{\lambda(\omega)^k} \sum_{\nu \leqslant \nu_0} a_\nu(\omega) \mu_\nu c_\nu, \tag{8}$$

where $\nu_0 = [\omega]$ and

$$a_\nu(\omega) = [\lambda(\omega) - \lambda_\nu]^k - [\lambda(\omega) - \lambda_{\nu+1}]^k \qquad \text{for } \nu < \nu_0,$$
$$a_{\nu_0}(\omega) = [\lambda(\omega) - \lambda_{\nu_0}]^k.$$

Introducing the partial sums C_n of the series $\sum c_n$ we can put (8) into the form

$$\sigma(\omega) = \frac{1}{\lambda(\omega)^k} \left\{ \sum_{\nu < \nu_0} [a_\nu(\omega)\mu_\nu - a_{\nu+1}(\omega)\mu_{\nu+1}] C_\nu + a_{\nu_0}(\omega)\mu_{\nu_0} C_{\nu_0} \right\}$$
$$= \sum_{\nu \leqslant \nu_0} b_\nu(\omega) C_\nu.$$

It is easy to see that $b_\nu(\omega) \to 0$ for $\omega \to \infty$ and that $\sum b_\nu(\omega) \to 0$. If we can show that

$$\sum_{\nu \leqslant \nu_0} |b_\nu(\omega)| = O(1), \tag{9}$$

we will have $\sigma(\omega) \to 0$ by a variant of the well-known Toeplitz theorem. To prove (9) observe that the last two terms in (9) are equal to $O(1)$, and therefore we may assume $\nu \leqslant \nu_0 - 2$. Putting

$$l_i = \lambda(\omega) - \lambda_{\nu+i} \qquad (i = 0, 1, 2),$$

we have then

$$a_{\nu+1}(\omega) - a_\nu(\omega) = -(l_0^k - 2l_1^k + l_2^k)$$
$$= -[k(k-1)(l_1 - l_0)^2 \xi^{k-2} - k(l_0 - 2l_1 + l_2)\eta^{k-1}]$$

where ξ and η are between l_0 and l_2. This shows that $a_\nu(\omega)$ increases with ν. Since also μ_ν has this property,

$$\sum_{\nu \leqslant \nu_0 - 2} |b_\nu(\omega)| = O(1) + \sum_{\nu \leqslant \nu_0} b_\nu(\omega) = O(1),$$

which completes the proof.

THEOREM 5. *Suppose that* λ_n *satisfies the conditions of Lemma* 1 *and that* (1) *is* $R(\lambda_n, 1)$*-summable to* $s(x)$ *on a set* E. *Then this series is even strongly* $R(\lambda_n, 1)$*-summable to* $s(x)$, *that is,*

$$R(\lambda_n, 1)\text{-lim } [s_\nu(x) - s(x)]^2 = 0 \qquad \text{a.e. on } E. \tag{10}$$

Proof. Using formula (8) for the $R(\lambda_n, 1)$ mean of (1), we have to show that

$$\sigma_n(x) = \frac{1}{\lambda(n)} \sum_{\nu \leqslant n} a_\nu(n)[s_\nu(x) - s(x)]^2 \to 0 \qquad \text{a.e. on } E. \tag{11}$$

First we prove

$$\sum \frac{\Delta\lambda_n}{\lambda_n}[s_n(x) - \sigma_n(x)]^2 < +\infty \qquad \text{a.e. on } E. \tag{12}$$

In fact, as in the proof of Theorem 2,

$$\begin{aligned} \sum_{n=1}^{\infty} \frac{\Delta\lambda_n}{\lambda_n} \int_a^b (s_n - \sigma_n)^2 dx &= \sum_{n=1}^{\infty} \frac{\Delta\lambda_n}{\lambda_n^3} \sum_{\nu \leqslant n} a_\nu^2 \lambda_\nu^2 \\ &= \sum_{\nu=1}^{\infty} a_\nu^2 \lambda_\nu^2 \sum_{n > \nu} \lambda_n^{-3} \Delta\lambda_n \leqslant \sum_\nu a_\nu^2 \lambda_\nu^2 \frac{1}{\lambda_\nu} \sum_{n \geqslant \nu} \left|\Delta\left(\frac{1}{\lambda_n}\right)\right| \\ &= A \sum a_\nu^2 < +\infty. \end{aligned}$$

Using Lemma 1, we obtain from (12),

$$R(\lambda_n, 1)\text{-lim } [s_n(x) - \sigma_n(x)]^2 = 0 \qquad \text{a.e. on } E.$$

By hypothesis and the regularity of $R(\lambda_n, 1)$,

$$R(\lambda_n, 1)\text{-lim } [\sigma_n(x) - s(x)]^2 = 0. \qquad \text{a.e. on } E,$$

and so (11) follows because $a_\nu(n) \geqslant 0$ and

$$(s_\nu - s)^2 \leqslant 2(s_\nu - \sigma_\nu)^2 - 2(\sigma_\nu - s)^2.$$

We now discuss the question of stability of convergence of s_{n_ν}, that is, the question of under which conditions $s_{n_\nu}(x) \to s(x)$ a.e. implies that, $s_{m_\nu}(x) \to s(x)$ a.e. for some other subsequence m_ν.

THEOREM 6. *Suppose that n_k is an increasing sequence of positive integers such that $n_{k+1} - n_k \to +\infty$ and that the partial sums $s_{n_k}(x)$ of the orthogonal series $\sum a_n\phi_n(x)$ are a.e. convergent. Then there is another sequence m_ν such that the number of the m_ν in the interval $I_k = [n_k, n_{k+1}]$ tends to $+\infty$, for which $s_{m_\nu}(x)$ is a.e. convergent.*

Proof. Let

$$s_{n_k}(x) \to s(x) \qquad \text{a.e.}$$

Choose $p_k \in I_k$ arbitrarily. Then $s_{p_k}(x)$ is a.e. convergent. In fact,

$$\int_a^b (s_{p_k} - s_{n_k})^2 dx = \sum_{\nu \in I_k} a_\nu^2$$

so that the integrals form a convergent series. Therefore

$$s_{p_k}(x) \to s(x) \qquad \text{a.e.}$$

Suppose now that the sequence

$$p_\nu = p_\nu^N$$

is so chosen that for any sufficiently large k there are exactly N of the p_ν in I_k. From the remark made above it follows again,

$$s_{p_\nu^N}(x) \to s(x) \qquad \text{a.e.}$$

For any fixed N and $k \to \infty$ we have

$$\rho_k^N(x) = \sup |s_{p_\nu^N}(x) - s(x)| \to 0 \qquad \text{a.e.,}$$

where the sup is taken for all $p_\nu^N \in I_l$ and $l \geqslant k$. Using a well-known theorem of Fréchet [3, p.561] we obtain integers $k_1 < k_2 < \ldots$ such that the "diagonal" subsequence

$$\rho_{k_N}^N(x) \to 0 \qquad \text{a.e.}$$

We define the sequence m_ν to coincide with the sequence p_ν^N in the intervals I_k, $k_N \leqslant k < k_{N+1}$, $N = 1, 2, \ldots$. Then

$$|s_{m_\nu}(x) - s(x)| \leqslant \rho_{k_N}^N(x),$$

and therefore

$$s_{m_\nu}(x) \to s(x) \qquad \text{a.e.,}$$

which completes the proof. (In a recent paper, Šneĭder [13] proves this result for the orthogonal system of Walsh. He uses special properties of this system.)

2. Some applications of theorems of § 1.

2.1 The most interesting application is to the existence of effective Riesz methods of summation which sum all orthogonal series of a certain class. (There is an erroneous statement contradicting this in [4, p.185].)

THEOREM 7. (i) *For any sequence* $\{a_n\}$ *with*

$$\sum a_n^2 < +\infty$$

there is a method $R(\lambda_n, 1)$ *which sums every orthogonal series* $\sum a_n\phi_n(x)$.
(ii) *For any orthogonal system* $\phi_n(x)$ *there is a method* $R(\lambda_n, 1)$ *effective for all series* $\sum a_n\phi_n(x)$.

Proof. (i) This follows from Theorem 4, if λ_n is taken such that (7) is fulfilled. (ii) By a theorem of Marcinkiewicz [9], to any system $\phi_n(x)$ there corresponds a sequence $n_1 < n_2 < \ldots$ such that the partial sums $s_{n_\nu}(x)$ of any series $\sum a_n\phi_n(x)$ with $\sum a_n^2 < +\infty$ are a.e. convergent. Putting $\lambda(n_\nu) = 2^\nu$ and extending λ by linear interpolation we obtain our result by Theorem 4.

2.2 THEOREM 8. *Let $\mu(n)$ be a positive function, increasing to infinity, and a_n a sequence such that, for any n, the number of integers m with the two properties $\mu(n) < \mu(m) < \mu(n) + 1$ and $a_m \neq 0$ is $O(\mu(n))$. If in addition*

$$\sum a_n^2 \log^2\mu(n) < +\infty, \tag{13}$$

then $\sum a_n\phi_n(x)$ is a.e. convergent.

Proof. Putting $\lambda(n) = \exp \mu(n)$, we deduce from (13) and Theorem 4 that $\sum a_n\phi_n(x)$ is a.e. $R(\lambda_n, 1)$-summable. Put

$$\delta_n(x) = \max_m \Big| \sum_{n<\nu<m} a_\nu\phi_\nu(x)\Big|,$$

for m subject to the condition $\lambda(n) < \lambda(m) < e\lambda(n)$ or to $\mu(n) < \mu(m) < \mu(n) + 1$. By a well-known lemma [4, p.162],

$$\int_a^b \delta_n^2 dx \leqslant A \log^2\mu(n) \sum_{\mu(n)<\mu(m)<\mu(n)+1} a_m^2.$$

Therefore

$$\sum \int \delta_n^2 dx < +\infty,$$

and $\delta_n(x) \to 0$. Using a Tauberian theorem of Szász [14] for the $R(\lambda_n, 1)$ method we obtain that $\sum a_n\phi_n(x)$ is a.e. convergent.

If $\mu(n)$ is increasing faster than $\log n$, this theorem is stronger than a theorem of Gál [2], which has the interval $n < m < 2n$ instead of that used in the text, and the additional hypothesis of C_1-summability of $\sum a_n\phi_n$. If $\mu(n)$ is increasing slower than $\log n$, both theorems are incomparable.

2.3 It is known that for any orthogonal system $\phi_n(x)$,

$$C_k\text{-}\lim \phi_n(x) = 0 \qquad \text{a.e., } k > \tfrac{1}{2} \tag{14}$$

(Lorentz [5]; for $k = 1$, this has also been proved by Banach [1]). We shall give a simple proof of this theorem, extending it at the same time to methods $R(\lambda_n,k)$.

THEOREM 9. *If $\Delta\lambda_n$, $\lambda_n/\Delta\lambda_n$ are increasing,* $0 < k \leqslant 1$, *and*

$$\sum_{n=1}^{\infty} (\Delta\lambda_n/\lambda_n)^{2k} \log^2 n < +\infty,$$

then

$$(15) \qquad R(\lambda_n, k)\text{-lim } \phi_n(x) = 0 \qquad a.e.$$

Proof. By the theorem of Menchoff-Rademacher we see that $\sum (\Delta\lambda_n/\lambda_n)^{2k}\phi_n(x)$ is a.e. convergent, and Lemma 1 gives the result.

The conditions are satisfied, for instance, if $\lambda_n = \exp n^a$, and $0 < a < 1 - (1/2k)$, or if $\lambda_n = n$ and $k > \frac{1}{2}$. This latter case gives (14). In the opposite direction we have: If

$$(16) \qquad \sum (\Delta\lambda_n/\lambda_{n+1})^{2k} = +\infty,$$

then there is an orthogonal system $\phi_n(x)$ such that (15) holds for no x. For (16) implies (Lorentz [5]) that there is a system $\phi_n(x)$ so that everywhere $\phi_n(x) \neq o(\lambda_{n+1}/\Delta\lambda_n)^k$, and by the well-known limitation theorem of $R(\lambda_n, k)$ we see that (15) cannot hold for any x.

3. Riesz means and general methods of summation. In this section we shall discuss relations among Riesz methods and general Toeplitz methods of summation. The results, which have some interest in themselves, will be applied in §4 to orthogonal series. Theorem 12 will not be used and was added for the sake of completeness.

Let A be a regular method of summation defined by the transformation

$$(17) \qquad S_m = \sum_{n=0}^{\infty} a_{mn}s_n \qquad (m = 0, 1, 2, \ldots).$$

$R(\lambda_n, 1)$ is again the method of arithmetic means (3) with λ_n increasing to $+\infty$. In what follows we will suppose that $\Delta\lambda_n$ increases and then we may even assume, without loss of generality, that $\Delta\lambda_n > 0$ for all n.

THEOREM 10. *Let $\Delta\lambda_n$ be increasing and $\Delta\lambda_n/\lambda_n$ decreasing to 0. In order that $R = R(\lambda_n, 1) \subset A$, it is necessary and sufficient that*

$$(18) \qquad \sum_{n=0}^{\infty} \frac{\lambda_n}{\Delta\lambda_n}|a_{mn} - a_{m\,n+1}| \leqslant M \qquad (m = 0, 1, \ldots)$$

for some constant M.

Proof. Expressing S_m in terms of σ_n by means of (3) we obtain

$$S_m = \sum_{n=0}^{\infty} (a_{mn}/\Delta\lambda_n)(\lambda_{n+1}\sigma_{n+1} - \lambda_n\sigma_n).$$

Since A is regular, the inclusion $R \subset A$ is equivalent to the condition that all S_m are defined and tend to 0 whenever $\sigma_n \to 0$. A necessary condition for this is

$$(19) \qquad a_{mn}\lambda_{n+1}/\Delta\lambda_n = O(1) \qquad \text{for any fixed } m.$$

Supposing that this is true, we split σ_n into two sequences, $\sigma_0, 0, \sigma_2, 0, \ldots$

and 0, σ_1, 0, $\sigma_3, \ldots$. Then our condition on S_m is equivalent to the same condition for the two sums

$$\sum_{n=0}^{\infty} \lambda_{2n+1}\left(\frac{a_{m,2n}}{\Delta\lambda_{2n}} - \frac{a_{m,2n+1}}{\Delta\lambda_{2n+1}}\right)\sigma_{2n+1},$$

$$\sum_{n=0}^{\infty} \lambda_{2n}\left(\frac{a_{m\,2n-1}}{\Delta\lambda_{2n-1}} - \frac{a_{m\,2n}}{\Delta\lambda_{2n}}\right)\sigma_{2n}. \tag{20}$$

By Toeplitz's theorem and using again the regularity of A, we see that both sums tend to 0 for any sequence $\sigma_n \to 0$ if and only if the norms of the lines in both expressions (20) are bounded. Since $\lambda_{n+1}/\lambda_n \to 1$, this is equivalent to one condition

$$\sum_{n=0}^{\infty} \lambda_n\left|\frac{a_{m\,n}}{\Delta\lambda_n} - \frac{a_{m\,n+1}}{\Delta\lambda_{n+1}}\right| \leqslant M \qquad (m = 0, 1, \ldots). \tag{21}$$

It follows that $R \subset A$ is true if and only if (19) and (21) hold and it remains to show that, under the hypotheses of the theorem, (19) and (21) are equivalent to (18). But the difference between the sums (18) and (21) does not exceed

$$\sum_{n=0}^{\infty} \frac{\lambda_n \Delta^2\lambda_n}{\Delta\lambda_n \Delta\lambda_{n+1}} |a_{m,n+1}|. \tag{22}$$

Here

$$\sum_n |a_{m\,n+1}|$$

is bounded, and the first factor of the sum (22) is bounded. In fact,

$$0 \leqslant \frac{\lambda_n \Delta^2 \lambda_n}{\Delta\lambda_n \Delta\lambda_{n+1}} \leqslant \frac{\lambda_n \Delta^2\lambda_n}{(\Delta\lambda_n)^2} = 1 + \frac{\lambda_n \Delta^2 \lambda_n - (\Delta\lambda_n)^2}{(\Delta\lambda_n)^2}$$

$$= 1 + \frac{\lambda_n \lambda_{n+1}}{(\Delta\lambda_n)^2}\Delta\left(\frac{\Delta\lambda_n}{\lambda_n}\right) \leqslant 1.$$

It follows that (18) and (21) are equivalent. On the other hand, (18) implies (19):

$$|a_{mn}|\lambda_n/\Delta\lambda_n \leqslant \sum_{\nu=n}^{\infty} \frac{\lambda_\nu}{\Delta\lambda_\nu}|a_{m\nu} - a_{m,\nu+1}| \leqslant M.$$

This proves the theorem.

THEOREM 11. *A regular method A contains a method $R(\lambda_n, 1)$ for which $\Delta\lambda_n/\lambda_n \to 0$ if and only if A is strongly regular, that is,*

$$\sum |a_{m,n} - a_{m,n+1}| \to 0 \qquad \textit{for } m \to \infty. \tag{23}$$

If this is true, there is even an $R(\lambda_n, 1) \subset A$ *with the properties described in Theorem* 10.

Proof. The condition is necessary. If $\Delta\lambda_n/\lambda_n \to 0$, then $R(\lambda_n, 1)$ is strongly regular (Lorentz [8]) and therefore A is strongly regular.

To show that the condition is sufficient, we have to find a sequence λ_n with properties described in Theorem 10 for which (18) holds. First, from (23) one deduces easily (compare [7]) the existence of a function $\Omega(n) > 0$ increasing to $+\infty$ such that

$$\sum \Omega(n) |a_{mn} - a_{m,n+1}| \to 0, \qquad m \cdot$$

We may obviously assume that $\Omega(n)$ has a continuous derivative $0 < \Omega'(n) < 1$. Put

$$\lambda(n) = \exp \int_0^n \Omega(u)^{-1} du.$$

Then $\lambda(n)$ is increasing; $\lambda'(n) = \lambda(n)\Omega(n)^{-1}$ and therefore $\lambda'(n)/\lambda(n)$ is decreasing to 0;

$$\lambda''(n) = \lambda(n)[1 - \Omega'(n)]\Omega(n)^{-2} \geqslant 0$$

and therefore $\lambda'(n)$ is increasing. From these facts we deduce that $\Delta\lambda_n = \lambda'(n + \theta)$ (θ here and later is some number $0 < \theta < 1$) is increasing with n; that

$$\log(1 + \Delta\lambda_n/\lambda_n) = \Delta(\log \lambda(n)) = \lambda'(n + \theta)/\lambda(n+\theta)$$

is decreasing to 0 and that therefore $\Delta\lambda_n/\lambda_n$ has the same property; and, finally, that

$$\lambda_n/\Delta\lambda_n = \lambda(n)/\lambda'(n + \theta) \leqslant \lambda(n)/\lambda'(n) = \Omega(n),$$

which implies (18). This proves our theorem.

The following theorem connects inclusion $R(\lambda_n, 1) \subset A$ with summability functions of the method A (see Lorentz [7], [8]).

THEOREM 12. *Let* $\Delta\lambda_n$ *be increasing,* $\Delta\lambda_n/\lambda_n$ *decreasing to* 0. *In order that all positive increasing functions of the form* $\Omega(n) = o(\lambda_n/\Delta\lambda_n)$ *should be summability functions of the second kind of regular method* A *it is necessary and sufficient that* $R(\lambda_n, 1) \subset A$.

Proof. By a result given elsewhere [7] the necessary and sufficient condition is

$$\sum_{n=0}^{\infty} \Omega(n) |a_{m,n} - a_{m,n+1}| \to 0 \tag{24}$$

for any such $\Omega(n)$. But an arbitrary function $\Omega(n) = o(\lambda_n/\Delta\lambda_n)$ has an increasing majorant which is also $o(\lambda_n/\Delta\lambda_n)$. This shows that (24) holds if and only if

$$\sum_{n=0}^{\infty} \epsilon_n\lambda_n/\Delta\lambda_n|a_{mn} - a_{m\,n+1}| \rightarrow 0$$

holds for any sequence $\epsilon_n \rightarrow 0$, $\epsilon_n \geqslant 0$. By regularity of A and Toeplitz's theorem this is in turn equivalent to (18). This proves the theorem.

4. Summation of orthogonal series by general methods.

4.1 We give first some general theorems. The most interesting fact (expressed in Theorems 13 and 14 below) is that a method (17) is, as far as we are concerned with summability of orthogonal series a.e., contained between two Riesz methods. More precisely we have

THEOREM 13. *For any regular method A there is a method $R(\lambda_n, 1)$ such that every orthogonal series which is a.e. A-summable is also a.e. $R(\lambda_n, 1)$-summable.*

THEOREM 14. *If A is strongly regular, there is an $R(\lambda_n, 1)$ method for which $\Delta\lambda_n$ is increasing and $\Delta\lambda_n/\lambda_n$ decreasing to 0 such that any orthogonal series which is a.e. $R(\lambda_n, 1)$-summable, is also a.e. A-summable.*

Proofs. Theorem 13 follows from Theorem 3 and the following result of Kaczmarz [4, p.183]: for any regular method A there is a sequence of integers $n_1 < n_2 < \dots$ such that $s_{n_\nu}(x)$ converges a.e. for any a.e. A-summable series $\sum a_n\phi_n(x)$. Theorem 14 follows from Theorem 11.

As an example, we note that from the remark after Theorem 4 and Theorem 13 it follows that for any method A there is a nowhere A-summable orthogonal series (Menchoff [10]).

The following theorems partly generalize some results of §1:

THEOREM 15. *For any strongly regular method A there is a sequence n_ν with $n_\nu/\nu \rightarrow \infty$ for $\nu \rightarrow \infty$ such that convergence a.e. of $s_{n_\nu}(x)$ implies A summability a.e. of $\sum a_n\phi_n(x)$.*

THEOREM 16. *If A is strongly regular, there is a function $\omega(n) = o\ (\log^2 n)$ such that*

$$\sum a_n^2\omega(n) < +\infty$$

implies A summability a.e. of the series $\sum a_n\phi_n(x)$.

Proofs. Theorem 15 is obtained from Theorems 14 and 3; we define the sequence n_ν in this case by $\lambda(n_\nu) = e^\nu$. Since $\Delta\lambda_n/\lambda_n \rightarrow 0$ implies $\Delta(\log\lambda_n) \rightarrow 0$, or $\log\lambda(n) = o(n)$, we have $\nu = o(n_\nu)$ or $n_\nu/\nu \rightarrow \infty$. Theorem 16 follows from Theorems 14 and 4 and the previous remarks.

4.2 We shall give simple proofs of two theorems of Menchoff, namely the following:

THEOREM 17. *Let* $A = (a_{mn})$ *be a regular method of summation with the property*

$$\epsilon_m = \max_n |a_{mn}| \to 0 \qquad \textit{for } m \to \infty, \tag{25}$$

and $\phi_n(x)$ *an orthogonal system. Then there is a rearrangement* $\psi_n(x)$ *of the* $\phi_n(x)$ *such that any orthogonal series* $\sum a_n\psi_n(x)$ *is a.e.* A*-summable.*

THEOREM 18. *If* $A^{(\nu)} = (a_{mn}^{(\nu)})$, $\nu = 1, 2, \ldots$ *is a sequence of regular methods with property* (25), *then for any orthogonal system* $\phi_n(x)$ *there is a rearrangement* $\psi_n(x)$ *such that any orthogonal series* $\sum a_n\psi_n(x)$ *is a.e.* $A^{(\nu)}$*-summable for any* ν.

Menchoff proved Theorem 17 in [**11**] and Theorem 18 (which contains the preceding one) in [**12**]. His proofs are very long. The proof given below differs from Menchoff's proof mainly in that we find (in Lemma 2) a kernel of the proof which is independent of orthogonal functions. Lemma 2 is a slight generalization of a known theorem (Lorentz [6, Theorem 8]).

LEMMA 2. *Let* A *be a regular method with the property* (25). *Then for any function* $\chi(n) > 0$ *increasing to* $+\infty$ *there is a function* $\Omega(n) > 0$, *increasing to* $+\infty$ *such that if the sequence of integers* $n_1 < n_2 < \ldots$ *has a counting function* $\leqslant \Omega(n)$, *then any sequence* s_n *with*

$$|s_{n_\nu}| \leqslant C\chi(\nu)$$

and $s_n = 0$ *for* $n \neq n_\nu$ *is* A*-summable to* 0.

Proof. Since

$$\sum_n |a_{mn}| < +\infty$$

for any m, there is a function $\chi_1(n) > 0$, increasing to $+\infty$, such that even

$$\sum_n |a_{mn}|\chi_1(n) < +\infty \qquad (m = 0, 1, 2, \ldots).$$

We can choose two increasing sequences p_m, q_m of positive integers such that $q_m < q_{m+1}$, $p_m \leqslant p_{m+1}$, $p_m \to \infty$, and for $m \to \infty$,

$$\sum_{n=q_m+1}^{\infty} |a_{mn}|\chi_1(n) = \eta_m \to 0, \qquad \epsilon_m p_m \chi(p_m) \to 0. \tag{26}$$

Let $\Omega_1(n)$ be a positive function increasing to $+\infty$ for which $\Omega_1(q_m) \leqslant p_m$. We also choose integers $N_1 < N_2 < \ldots$ such that

$\chi(\nu) \leqslant \chi_1(N_\nu)$; if $\Omega_2(n)$ is the counting function of the N_ν, and a sequence n_ν has the counting function $\leqslant \Omega_2(n)$, then $n_\nu \geqslant N_\nu$ and therefore

$$(27) \qquad \chi(\nu) \leqslant \chi_1(n_\nu) \qquad (\nu = 1, 2, \ldots).$$

We take $\Omega(n) = \min\,(\Omega_1(n), \Omega_2(n))$.

Suppose now that the sequences n_ν and s_n have the properties described in the lemma. Then

$$\begin{aligned} |\sigma_m| = |\sum_{n=0}^{\infty} a_{mn}s_n| &\leqslant \sum_{n=0}^{q_m} |a_{mn}s_n| + \sum_{q_m+1}^{\infty} |a_{mn}s_n \\ &\leqslant \epsilon_m C \sum_{\nu=1}^{p_m} \chi(\nu) + C\eta_m \\ &\leqslant C\epsilon_m p_m \chi(p_m) + C\eta_m \to 0, \end{aligned}$$

by (26) and (27). This proves our lemma.

Corollary. If we assume in Lemma 2 that $s_n \to s$ for $n \neq n_\nu$, $n \to \infty$, without changing the other parts of the hypothesis, then we see that the sequence s_n is A-summable to s.

Proof of Theorem 17. Using the theorem of Marcinkiewicz quoted in §2.1, we obtain a sequence $p_1 < p_2 < \ldots$ so that, in particular, any orthogonal series $\sum a_\nu \phi_{p_\nu}(x)$ is a.e. convergent. Let $q_\nu = q(\nu)$ be all positive integers not contained in the sequence p_ν. Again by Marcinkiewicz's theorem there is a subsequence $q(\nu(k))$ such that the partial sums $s_{\nu(k)}(x)$ of any orthogonal series $\sum a_\nu \phi_{q_\nu}(x)$ are a.e. convergent. Let $\Omega(n)$ be defined by Lemma 2 for the method A and let $\chi(\nu) = \nu^2$. We then choose a sequence of intervals I_k, $r_k \leqslant n < r'_k$, so that the length of I_k is exactly $r'_k - r_k = \nu(k) - \nu(k-1)$ and that the sequence of all integers n_ν, belonging to these I_k, has the counting function $\leqslant \Omega(n)$. Let m_ν denote all positive integers not belonging to the sequence n_ν. We define a rearrangement $\psi_1, \psi_2, \ldots$ of the $\phi_1, \phi_2, \ldots$ taking

$$\psi_{n_\nu} = \phi_{q_\nu}, \qquad \psi_{m_\nu} = \phi_{p_\nu} \qquad (\nu = 1, 2, \ldots).$$

We write $\sum a_n\psi_n = \sum a'_n\psi_n + \sum a''_n\psi_n$, where $a'_n = a_n$ if n is an m_ν and $a'_n = 0$ if n is an n_ν, and $a''_n = a_n - a'_n$. Then $\sum a'_n\psi_n$ is a.e. A-summable, since this series reduces, after omission of vanishing terms, to $\sum a_{m_\nu}\phi_{p_\nu}$ and this is a.e. convergent. The nth partial sum of $\sum a''_n\psi_n$ is constant and equal to $s_{\nu(k)}(x)$, if n varies between r'_{k-1} and r_k, and this has a.e. a limit $s(x)$ for $k \to \infty$. On the other hand it is an elementary fact that the νth partial sum of an orthogonal series is $O(\nu^2)$ a.e. Applying this to $\sum a_{n_\nu}\phi_{q_\nu}$, to which $\sum a''_n\psi_n$ reduces if vanishing terms are omitted, we obtain, by the corollary to Lemma 2, that $\sum a''_n\psi_n$ is a.e. A-summable to $s(x)$. This completes the proof.

Proof of Theorem 18. This may be obtained by making minor changes in the preceding proof. But Theorem 18 is also an easy corollary of Theorem 17. Without loss of generality, we may assume that $|a_{mn}^{(\nu)}| \leqslant \nu^{-1}$ for all m, n, and ν (otherwise we would replace $A^{(\nu)}$ by an equivalent method, multiplying some of the first lines of $A^{(\nu)}$ by a small constant). If $\mathrm{l}_m^{(\nu)}$ is the mth line of $A^{(\nu)}$, we consider the method A with the lines $\mathrm{l}_1^{(1)}, \mathrm{l}_1^{(2)}, \mathrm{l}_2^{(1)}, \mathrm{l}_1^{(3)}, \mathrm{l}_2^{(2)}, \ldots$. Then Theorem 18 follows by applying Theorem 17 to the method A.

REFERENCES

1. S. BANACH, *Sur la valeure moyenne des fonctions orthogonales*, Bull. Acad. Sci. Cracovie, A (1919), 67-72.
2. S. GÁL, *Sur les séries orthogonales C_1-summables et $\lambda(n)$-lacunaires*, Comptes Rendus, Paris, vol. 227 (1948), 1140–2.
3. H. HAHN, *Theorie der reellen Funktionen*, Bd. 1 (Berlin, 1921).
4. S. KACZMARZ and H. STEINHAUS, *Theorie der Orthogonalreihen* (Warsaw-Lwow, 1935).
5. G. G. LORENTZ, *Über die Mittelwerte der Funktionen eines Orthogonalsystems*, Math. Zeit., vol. 49 (1944), 724–33.
6. ———, *A contribution to the theory of divergent sequences*, Acta Math., vol. 80 (1948), 167–90.

7, 8. ———, *Direct theorems on methods of summability I, II*, Can. J. Math., vol. 1 (1949), 305–19; vol 3 (1951), 236–56.

9. J. MARCINKIEWICZ, *Sur la convergence des séries orthogonales*, Studia Math., vol. 6 (1936), 39–45.
10. D. MENCHOFF, *Sur les séries de fonctions orthogonales II*, Fund. Math., vol. 8 (1926), 56–108.
11. ———, *Sur la sommation des séries de fonctions orthogonales par des méthodes linéaires*, Bull. Acad. Sci. URSS, Ser. math. (1937), 203–29.
12. ———, *Sur la sommation des séries de fonctions orthogonales par des méthodes de Cesàro*, Mat. Sbornik, n.s., vol. 8 (1940), 121–36.
13. A. A. ŠNEĬDER, *On the convergence of subsequences of the partial sums of Fourier series of Walsh functions*, Doklady Akad. Nauk SSSR, n.s., vol. 70 (1950), 969-71.
14. O. SZÁSZ, *Converse theorems for summability for Dirichlet's series*, Trans. Amer. Math. Soc., vol. 39 (1936), 117-30.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLV : SERIES III : JUNE, 1951
SECTION THREE

Means of Fourier Constants*

G. M. PETERSEN

Presented by G. deB. ROBINSON, F.R.S.C.

THE spaces $\Lambda(\alpha)$, $M(\alpha)$, $\Lambda(\alpha, p)$ have been introduced by G. Lorentz [4], and the space $\Lambda^*(\alpha, p)$ by Lorentz and Halperin [5].

A real measurable function $f(x)$, $0 < x < 1$, belongs to $\Lambda(\alpha)$, $0 < \alpha \leqslant 1$, if and only if

$$||f||_{\Lambda(\alpha)} = ||f|| = \alpha\int_0^1 x^{\alpha-1}f^*(x)dx$$

is finite. Here, and in the sequel, we denote by $f^*(x)$ the decreasing rearrangement of $|f(x)|$ (see Hardy, Littlewood, and Pólya [3]).

A measurable function $f(x)$ belongs to $M(\alpha)$, $0 < \alpha < 1$, if

$$||f||_{M(\alpha)} = ||f|| = \sup_e \left\{(me)^{-\alpha}\int_e |f(x)|dx\right\}$$

is finite. Here e is an arbitrary measurable subset of (0, 1). Introducing $f + g$ and af in the natural way, we can show that these spaces are normed and complete and hence Banach spaces (see [4]).

A real measurable function $f(x)$, $0 < x < 1$, belongs to $\Lambda(\alpha, p)$, $0 < \alpha \leqslant 1$; $p > 1$, if

$$||f||_{\Lambda(\alpha,p)} = ||f|| = \left\{\int_0^1 x^{\alpha-1}f^*(x)^p dx\right\}^{1/p}$$

is finite.

We shall write $g_1(x) \succ g_2(x)$ if

$$\int_a^x g_1(t)dt \geqslant \int_a g_2(t)dt$$

for all $a \leqslant x \leqslant b$. A measurable function $f(x)$ belongs to $\Lambda^*(\alpha,p)$, $0 < \alpha < 1$, if there exists a $D(x)$ positive and decreasing, and such that

$$f^*(x) \prec x^{\alpha-1}D(x),$$

and

$$\frac{1}{q} + \frac{1}{p} = 1,$$

---

*Part of a thesis written in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Toronto.

$$||f||_{\Lambda^*(\alpha\ p)} = ||f|| = \inf_{x^{\alpha-1}D \succ f^*} \left\{\int_0^1 x^{\alpha-1} D^q(x)dx\right\}^{1/q} < +\infty$$

We state the following:

THEOREM 1. If $a_1, a_2, \ldots, a_n, \ldots$ *are the Fourier constants of a function*

$$f(x) \sim \sum_{n=1}^{\infty} a_n \cos nx$$

of any one of the spaces $\Lambda(\alpha)$, $M(\alpha)$, $\Lambda(\alpha, p)$, $\Lambda^*(\alpha, p)$, *then* $A_1, A_2, \ldots, A_n, \ldots$ *are the Fourier constants of a function of the same space, where*

$$A_n = n^{-1}\sum_1^n a_k.$$

This theorem was proved for the spaces L^p $(p \geqslant 1)$ by Hardy [2]; we shall indicate where the proof of Theorem 1 is different from that for the case L^p. As in Hardy's proof, the statement may be reduced to the assertion that the transformation

$$h(x) = \int_x^{\pi} \frac{f(u)}{u} du$$

maps any of the spaces $\Lambda(\alpha)$, $M(\alpha)$, $\Lambda(\alpha, p)$, $\Lambda^*(\alpha, p)$ into itself.

In the case of the spaces $M(\alpha)$, $\Lambda(\alpha, p)$, and $\Lambda^*(\alpha, p)$ this can be done by first showing that if $k(x)$ belongs to $\Lambda(\alpha)$, $\Lambda^*(\alpha, p)$, or $\Lambda(\alpha, p)$ so does $K(x)$, defined by

$$K(x) = \frac{1}{x}\int_0^x k^*(u)du.$$

It then follows that

$$\begin{aligned}\int_0^{\pi} |h|\,|k| dx &\leqslant \int_0^{\pi} |k(x)| dx \int_x^{\pi} \frac{|f(u)|}{u} du = \int_0^{\pi} \frac{|f(u)|}{u} du \int_0^u |k(x)| dx \\ &\leqslant \int_0^{\pi} \frac{|f(u)|}{u} du \int_0^u k^*(x) dx = \int_0^{\pi} |f(u)| K(u) du \\ &\leqslant ||f||_X \, ||K||_Y\end{aligned}$$

exists for every k belonging to Y and hence h belongs to X; where X is the space $M(\alpha)$, $\Lambda(\alpha, p)$, or $\Lambda^*(\alpha, p)$, and Y is $\Lambda(\alpha)$, $\Lambda^*(\alpha, p)$, or $\Lambda(\alpha, p)$ respectively. (See [4] and [5].)

This procedure does not work if $h \in \Lambda(\alpha)$, but there is a direct proof:

$$h^*(x) \leqslant \int_x^{\pi} \frac{f^*(t-x)}{t} dt = \int_0^{\pi - x} \frac{f^*(u)}{x+u} du,$$

$$\begin{aligned}||h(x)||_{\Lambda(\alpha)} &= \alpha \int_0^{\pi} x^{\alpha-1} h^*(x) dx \leqslant \alpha \int_0^{\pi} x^{\alpha-1} dx \int_0^{\pi} \frac{f^*(u)du}{x+u} \\ &= \alpha \int_0^{\pi} f^*(u)du \int_0^{\pi} \frac{x^{\alpha-1}dx}{x+u}.\end{aligned}$$

However, making the substitution $x = ut$,

$$\int_0^{\pi} \frac{x^{a-1}dx}{x+u} \leqslant u^{a-1}\int_0^{\infty} \frac{t^{a-1}dt}{1+t} = Cu^{a-1}$$

Substituting back,

$$a\int_0^{\pi} x^{a-1}h^*(x)dx \leqslant aC\int_0^{\pi} u^{a-1}f^*(u)du < +\infty,$$

since f belongs to $\Lambda(a)$. Hence $h(x)$ belongs to $\Lambda(a)$.

We next state a dual to Theorem 1.

THEOREM 2. *If* $b_1, b_2, \ldots, b_n, \ldots$ *are the Fourier constants of a function*

$$g(x) \sim \sum_1^{\infty} b_n \cos nx$$

of one of the spaces $\Lambda(a)$, $M(a)$, $\Lambda(a, p)$, $\Lambda^*(a, p)$ $(0 < a < 1)$, *then* $B_1, B_2, \ldots, B_n, \ldots$ *are also the Fourier constants of a function of the same space, where*

$$B_n = \sum_{k=n}^{\infty} k^{-1}b_k.$$

The proof of Theorem 2 is similar to that for the case L^p given by Bellman [1]. All that is required is an estimate for

$$|\sum_{\lambda=0}^{n} b_k|,$$

where b_k is the Fourier constant of a function in $\Lambda(a)$, $M(a)$, $\Lambda(a, p)$, or $\Lambda^*(a, p)$. $\Lambda(a)$ is contained in $L^{a^{-1}}$ and $\Lambda(a, p)$ is contained in $L^{p/a}$. The estimate of this sum for L^p has been made by Bellman [1]. However, $\Lambda^*(a, p)$ is contained in $M(a/p)$ so we need only the estimate given in the following

LEMMA. *If*

$$H(x) \sim \sum_1^{\infty} b_n \cos nx$$

is a function of $M(a)$, $0 < a < 1$, *then*

$$|\sum_1^N b_n| = O(N^{1-a}).$$

Proof.

$$\sum_1^N b_n = \frac{1}{\pi}\int_0^{\pi} H(x)\left[\sum_1^N \cos nx\right]dx.$$

Using the characteristic "Hölder type" inequality for $M(a)$ spaces (see [4, p. 41]) we obtain

$$|\sum_1^N b_n| \leqslant ||H||_{M(\alpha)} \cdot \alpha \int_0^\pi x^{\alpha-1}\left\{\sum_1^N \cos nx\right\}^* dx,$$

since $H(x)$ belongs to $M(\alpha)$. We break

$$\sum_1^N \cos nx$$

into two parts:

$$F_1(x) = \begin{cases} \displaystyle\sum_1^N \cos nx & 0 \leqslant x \leqslant \dfrac{\pi}{2N} \\ 0 & \dfrac{\pi}{2N} \leqslant x \leqslant \pi, \end{cases}$$

$$F_2(x) = \begin{cases} 0 & 0 \leqslant x \leqslant \dfrac{\pi}{2N} \\ \displaystyle\sum_1^N \cos nx & \dfrac{\pi}{2N} \leqslant x \leqslant \pi. \end{cases}$$

Hence

$$\int_0^\pi x^{\alpha-1}\left\{\sum_1^N \cos nx\right\}^* dx \leqslant \int_0^\pi x^{\alpha-1}F_1^*(x)dx + \int_0^\pi x^{\alpha-1}F_2^*(x)dx.$$

Clearly $F_1(x) \leqslant N$ in $0 < x \leqslant \pi/2N$, so that

$$\int_0^\pi x^{\alpha-1}F_1^*(x)dx \leqslant N\int_0^{\pi/2N} x^{\alpha-1}dx = C_1N^{1-\alpha}.$$

$$F_2(x) = \frac{\sin \frac{1}{2}(2N+1)x}{2 \sin \frac{1}{2}x} \quad \text{in} \quad \frac{\pi}{2N} < x \leqslant \pi,$$

hence

$$\int_0^\pi x^{\alpha-1}F_2^*(x)dx = \int_0^{\pi-\pi/2N} x^{\alpha-1}\left\{\frac{\sin \frac{1}{2}(2N+1)(x+\pi/2N)}{2 \sin \frac{1}{2}(x+\pi/2N)}\right\}^* dx.$$

Now $(fg)^* \prec f^*g^*$ and, moreover, $g_1 \prec g_2$ implies

$$\int_a^b x^{\alpha-1}g_1dx \leqslant \int_a^b x^{\alpha-1}g_2dx$$

(or, more generally, $\int hg_1 \leqslant \int hg_2$ if h is positive and decreasing). Therefore

$$\int_0^\pi x^{\alpha-1}F_2^*(x)dx \leqslant \frac{1}{2}\int_0^{\pi-\pi/2N} x^{\alpha-1}\{(\sin \tfrac{1}{2}(x+\pi/2N))^{-1}\}^* \cdot\{\sin \tfrac{1}{2}(2N+1)(x+\pi/2N)\}^*dx,$$

and since

$$\{\sin \tfrac{1}{2}(2N+1)(x+\pi/2N)\}^* \leqslant \cos \tfrac{1}{2}x,$$

$$\sin x > \frac{2}{\pi}x,$$

the integral does not exceed

$$\frac{1}{2}\int_0^{\pi-\pi/2N}\frac{x^{\alpha-1}\cos\frac{1}{2}x}{\sin\frac{1}{2}(x+\pi/2N)}dx \leqslant \frac{\pi}{2}\int_0^{\pi-\pi/2N}\frac{x^{\alpha-1}}{x+\pi/2N}dx.$$

Now

$$\int_0^{\pi}\frac{x^{\alpha-1}}{x+\delta}dx = O(\delta^{\alpha-1}) \qquad (\delta\to 0),$$

as is seen by making in the integral the substitution $x = \delta t$. Hence

$$\int_0^{\pi} x^{\alpha-1}F_2^*(x)dx \leqslant C_2N^{1-\alpha}.$$

Altogether then,

$$\left|\sum_1^N b_n\right| = O(N^{1-\alpha}).$$

We have operated on the Fourier coefficients with the $(c, 1)$ matrix. It is interesting to examine the result of operating on Fourier constants with the (c, α) matrix. In the following three theorems, we denote by A_n^α the quantity

$$\binom{n+\alpha}{n} \simeq \frac{n^\alpha}{\Gamma(\alpha+1)}.$$

THEOREM 3. *If $a_1, a_2, \ldots, a_n, \ldots$ are the Fourier constants of an even function*

$$f(x) \sim \sum_1^\infty a_n \cos nx$$

belonging to L^p ($p > 2$; $\alpha + 1/p > 1$), then $B_1, B_2, \ldots, B_n, \ldots$, where

$$B_n = \frac{1}{A_n^\alpha}\sum_1^n A_{n-k}^{\alpha-1}a_k,$$

are the Fourier constants of an even function $g(x)$ belonging to $L^{p'}$ ($p' < p$).

THEOREM 4. *If $a_1, a_2, \ldots, a_n, \ldots$ are the Fourier constants of an even function*

$$f(x) \sim \sum_1^\infty a_n \cos nx$$

belonging to L^p ($p > 2$; $\alpha + 1/p > 1$), then $B_1, B_2, \ldots, B_n, \ldots$, where

$$B_n = \frac{1}{A_n^\alpha}\sum_1^n A_k^{\alpha-1}a_k,$$

are the Fourier constants of an even function $g(x)$ of $L^{p'}$ ($p' < p$).

THEOREM 5. *If $a_1, a_2, \ldots, a_n, \ldots$ are the Fourier constants of a function in L^2 and $\alpha > \frac{1}{2}$ then*

$$A_n = \frac{A_1^{\alpha-1}a_1 + A_2^{\alpha-1}a_2 + \ldots + A_n^{\alpha-1}a_n}{A_n^{\alpha}}$$

is the nth coefficient of a function in L^2.

In the proof of Theorem 3 and of Theorem 4, it is shown from the integral representation of B_n that $|B_n| < Cn^{-1/q}$. The result then follows from the Hausdorff-Young theorem.

In the proof of Theorem 5, we use an inequality of Hardy, namely,

$$\sum \lambda_n \left(\frac{B_n}{\Lambda_n}\right)^2 \leqslant 4 \sum \lambda_n b_n^2,$$

where

$$B_n = \lambda_1 b_1 + \lambda_2 b_2 + \ldots + \lambda_n b_n,$$
$$\Lambda = \lambda_1 + \lambda_2 + \ldots + \lambda_n,$$

and we suppose that $b_n > 0$, $\lambda_m > 0$, and that $\sum \lambda_n b_n^2$ is convergent. We set

$$b_n = \frac{1}{\sqrt{\lambda_n}} a_n$$

and the result follows from the Riesz-Fischer theorem.

In conclusion, I want to express my gratitude to Dr. G. Lorentz of the University of Toronto for his valuable help and criticism in the preparation of this paper.

REFERENCES

1. R. Bellman, *A note on a theorem of Hardy on Fourier constants*, Bull. Amer. Math. Soc., vol. 40 (1944), 741-44.
2. G. H. Hardy, *Notes on some points in the integral calculus*, Note LXVI, Messenger of Math., vol. 58 (1928), 50-2.
3. G. H. Hardy, L. E. Littlewood, and G. Pólya, *Inequalities* (Cambridge University Press, 1934).
4. G. G. Lorentz, *Some new functional spaces*, Ann. of Math., vol. 51 (1950), 37-55.
5. Some unpublished papers of G. G. Lorentz and I. Halperin.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLV : SERIES III : JUNE, 1951
SECTION III

The Dielectric Constants of Binary Mixtures*

J. W. TOMECKO and W. H. HATCHER, F.R.S.C.

INTRODUCTION

IN the course of investigations into the nature of organic peroxides it became necessary to ascertain what physical changes occur in non-aqueous solutions, mainly with respect to polarity, in the presence of a carbonyl group. Acetone was chosen because of its solubility in both water and many carbon compounds and because previous researches permitted of a close check on the methods here employed.

The dielectric properties of mixtures of fatty acids with non-polar solvents have been studied previously using solvents of low dielectric constant. Thwing (1893) attempted to measure the entire range of binary mixtures of acetic acid and water. He encountered considerable difficulty in making capacitance measurements because of conductance. Acetone has a moderately high dielectric constant and reasonably small conductance. A parallel between acetone and water as solvents was demonstrated by Kateswaram (1938), who found a shift in the Raman spectrum when acetic acid is dissolved in acetone, comparable to that found when the acid is dissolved in water.

APPARATUS AND METHOD

The apparatus and method used involved a substitution method with two oscillators operating at 730 kilocycles per second. One was used as reference standard and the other as working oscillator. "Equality" was determined by the heterodynebeat method. A circuit diagram is given in Fig. 1 where small mutual coupling exists between coils L_3 and L_5 and between L_4 and L_6. This apparatus and method are similar to those used by Smyth (1931), Williams (1930), and others but certain modifications were made to minimize the effect of leakage conductance present in the test samples.

---

*Presented as part fulfilment of the requirements for the degree of Ph.D., McGill University. The authors wish to express their appreciation to Professors Marshall, Pounder, and Watson for their invaluable assistance. This work was carried on in the Chemical Laboratories of Loyola College in Montreal.

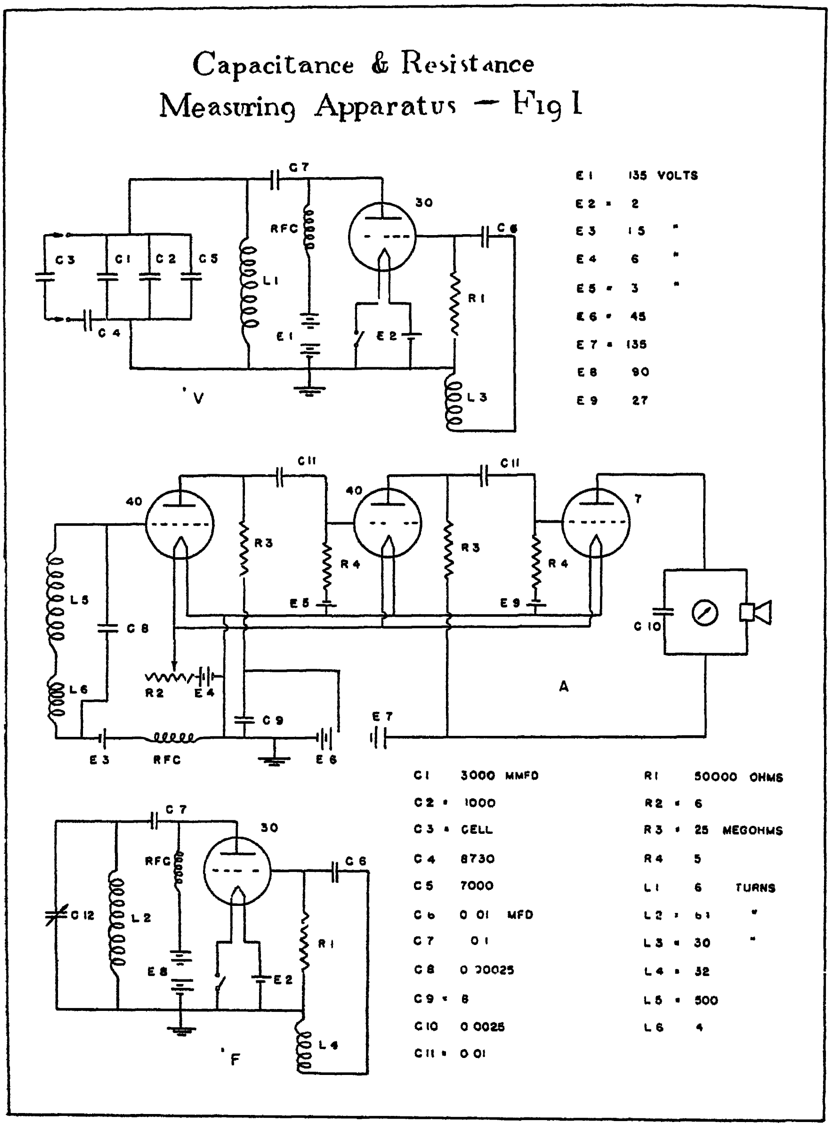
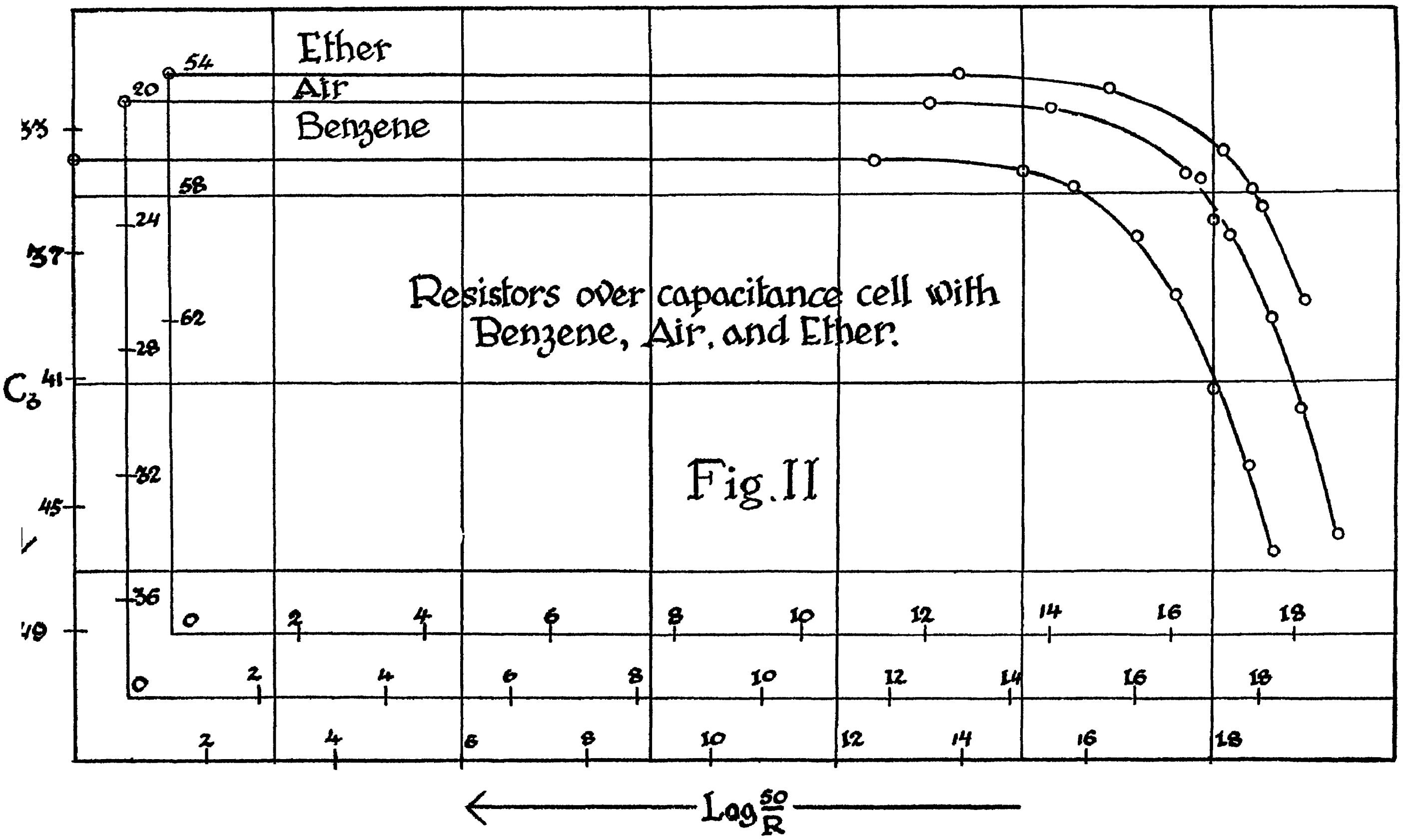

Capacitance & Resistance
Measuring Apparatus — Fig 1
C7
30
RFC
C6
C3
C1
C2
C5
L1
R1
E1
E2
C4
V
L3
E1 135 VOLTS
E2 = 2
E3 1.5 "
E4 6 "
E5 = 3 "
E6 = 45
E7 = 135
E8 90
E9 27
C11
C11
40
40
7
R3
R4
R3
R4
L5
E5
E9
C8
C10
R2
E4
A
L6
C9
E7
E3
RFC
E6
C1 3000 MMFD
C2 = 1000
C3 = CELL
C4 8730
C5 7000
C6 0.01 MFD
C7 0.1
C8 0.00025
C9 = 8
C10 0.0025
C11 = 0.01
R1 50000 OHMS
R2 = 6
R3 = 25 MEGOHMS
R4 5
L1 6 TURNS
L2 = 61 "
L3 = 30 "
L4 = 32
L5 = 500
L6 4
C7
30
RFC
C6
C12
L2
E8
E2
R1
L4
F

FIGURE 1



Resistors over capacitance cell with Benzene, Air, and Ether.

Fig. II

TABLE I

DIELECTRIC CONSTANTS OF
ACETONE, TOLUENE, PROPIONALDEHYDE AND ISOBUTYRIC, BUTYRIC, PROPIONIC AND ACETIC ACIDS

Material	C_s	C_L	C_O	E	*Other Workers*
Acetone	218.50	9.45	10.55	19.8	See Table II
Acetone	495.30	20.40	23.9	19.8	See Table II
Toluene	76.90	20.20	23.9	2.37	2.31 (18°C.)—Schupp 2.36 (i)—Grutzmacher 2.37 (i)—Williams (1926)
Propionaldehyde	437.30	20.60	23.9	17.4	18.9—4 (17°C.)—Drude (1897) 14.4 (14°C.)—Thwing
Isobutyric acid	36.70	9.45	10.55	2.58	2.68 (20°C.)—Drude (1898) 2.717 (i) (43)—Piekara
Butyric acid	89.20	20.10	23.9	2.90	2.966 (i) (51)—Smyth (1930) 2.85 (20°C.)—Int. Critical Tables 3.006 (1) (43)—Piekara
Propionic acid	96.60	20.10	23.9	3.21	3.19 (19°C.)—Int. Critical Tables 3.316 (i)— Piekara
Acetic acid	167.50	21.35	23.9	6.14	6.18 (i)—Smyth (1930) 6.476 (i)—Piekara

The principle of the modifications is to use a large C/L ratio for the oscillating circuit of the test oscillator thus reducing the damping effect of the leakage conductance. The validity of this principle was verified experimentally by connecting non-inductive resistances across the test cell when filled with a liquid of low leakage conductance. The resulting effect on the apparent C was then measured. The effect is shown in Fig. 2. In this way the cell conductance was shown to be so small that no error would result from it in all cases.

DIELECTRIC CONSTANTS OF PURE SUBSTANCES

Dielectric constants of seven pure substances were redetermined and our values found to be in good agreement with those of other workers (Table I). The nature of the leak effect is demonstrated by comparing the present experiment with previous ones, particularly that of Akerlof (1932) in the case of acetone (Table II).

TABLE II

DIELECTRIC CONSTANT OF ACETONE

Author	*Method*	F(KC)	(M)	*Temp.* (C)	E
Drude (1897)	Aulio			19°	20.7
Walden (1903)	Bridge			25°(i)	19.95
Int. Critical Tables	Bridge			20°	21.3 ±1
Grattunder (1930–1)	Heterodyne		450–500	25°	20.87
Kerr (1926)	Heterodyne	3,000		17°	20.89
Lunt & Rau (1930)	Bridge	42		25°	20.2
Lunt & Rau	Bridge	872		25°	20.6
Lunt & Rau	Bridge	883		25°	20.4
Albright (1937)				25°	20.74
L.B.R.S.P.C.T.	Resonance		81–84.5	25°	20.88
Akerlof (1932)	Resonance	2,000	150	25°	19.1
This work	Heterodyne	730	410	25°	19.8

Two types of error may result from the leak effect. These may be denoted as positive and negative effects. The leak effect is positive when the cell condenser capacitance is greater than the capacitance remaining in the standard condenser after the cell condenser has been placed in the circuit and the oscillator has been brought to its original frequency. Capacitance and dielectric values in such cases where conductance is operative will be too high. Smyth observes that capacitance errors, as high as 100 per cent, may result with methyl alcohol when conductance occurs. A negative leak effect results when the capacitance of the standard condenser is greater than that of the cell even when the condenser cell has been placed in circuit. In this case, as the material in the condenser cell becomes conductive, the leak effect causes the capacitance of the standard condenser to increase more than that of the cell condenser resulting in low capacitance and dielectric constant values.

The dielectric constants of acetone, as determined by Akerlof are considerably lower than those of other authors (Table II). In the tank circuit, Akerlof placed the standard condenser parallel to the cell condenser which consists of two 5 cm. discs 2 mm. apart, having a capacitance of 10.0 uufd. with air, or approximately 200 uufd. with acetone. The standard condenser had a capacitance of 1500 uufd. which after tuning the circuit would leave 1300 uufd. in the standard condenser as against 200 uufd. in the cell. This condenser arrangement will produce, with conductance, capacitance and dielectric values which are too low.

The dielectric constants as found in this work, when neither negative nor positive leak effects due to conductance were operative, are higher than the value obtained by Akerlof and slightly lower than those given by most authors (Table II).

DIELECTRIC CONSTANTS OF BINARY MIXTURES

Measurements were made on binary mixtures of acetone with benzene, with isobutyric, butyric, propionic, and acetic acids, with propionaldehyde and with toluene. The mixtures were made gravimetrically in 10, 20, 30, 40, 50, 60, 70, 80, and 90 per cent. The values determined for the dielectric constants are shown graphically in Fig. 3. The polarization values calculated from the relationship

$$P_{1,2} = \frac{E-1}{E+2} \cdot \frac{f_1M_1 + f_2}{d}$$

are shown similarly in Fig. 4.

REAGENTS

All reagents were fractionated with an eight-inch Hempel column packed with ¼″ glass helices about 1 mm. thick.

Merck's Reagent Quality Benzene, thiophene-free, was allowed to stand over sodium until effervescence ceased. It was then fractionated and the portion boiling at 80.2° C. (Corr.) was collected and used. The ether was dried over anhydrous calcium chloride, treated with sulphuric acid, and then distilled and stored over sodium wire. It was freshly distilled from sodium when required.

Acetone, Merck's Reagent Quality, was fractionated after dissolving a small amount of potassium permanganate and adding a few drops of concentrated sulphuric acid. The portion boiling at 56.4° C. (Corr.) was used. This method of purification gave acetone dielectric values about 2 per cent lower than those obtained when acetone was fractionated after having been refluxed over freshly burned lime.

Eastman isobutyric and butyric acids were fractionated. The portions boiling between 153.5° and 154° C. (Corr.), and 162.0° and 162.7° C. (Corr.) respectively, were used.

Th. Schuchardt propionic acid was fractionated twice. The portion boiling between 140.2° and 140.7° C. (Corr.) was used.

General Chemical 99.5 per cent glacial acetic acid was fractionated twice and the fraction between 117.6° and 117.8° C. (Corr.) was used. Smyth and Rogers (51) made twelve crystallizations of glacial acetic acid which boiled at 117.8° C. Because of the hydroscopic nature of acetic acid, purification by crystallization was unsatisfactory.

Elmer and Amend propionaldehyde was fractionated and the fraction boiling between 48° and 49.0° C. (Corr.) was collected and used.

Toluene, Merck's Reagent Quality, was placed over sodium until effervescence ceased. The portion distilling at 110.5° C. (Corr.) was used.

Discussion of Results

In Fig. 3 are seen three kinds of curves; the linear, acetone-propionaldehyde; the convex, acetone-acetic and propionic acids; and the concave. In Fig. 4 the polarizations contain one linear (acetone-propionaldehyde) and five convex curves; that for acetone-toluene alone shows a maximum at 70 per cent acetone content.

The dielectric constants and polarization curves of mixtures of two components which are both polar have not been studied as extensively as the mixtures where one of the components is non-polar since the relationships are too complicated to interpret (Debye). Thwing measured the dielectric constants of acetic acid-water mixtures obtaining a curve which was, in general, convex. The dielectric constants of acetone-water mixtures determined by Akerlof gave a concave curve but the polarization composition relationship was linear. Kerr (1926) and Graffunder (1931) both found that the dielectric constants of binary mixtures of chloroform-acetone gave a convex curve in place of the concave type which is the general rule for a polar and molecularly associated material in a non-polar medium. The nitrobenzene-acetone dielectric constant curve obtained by

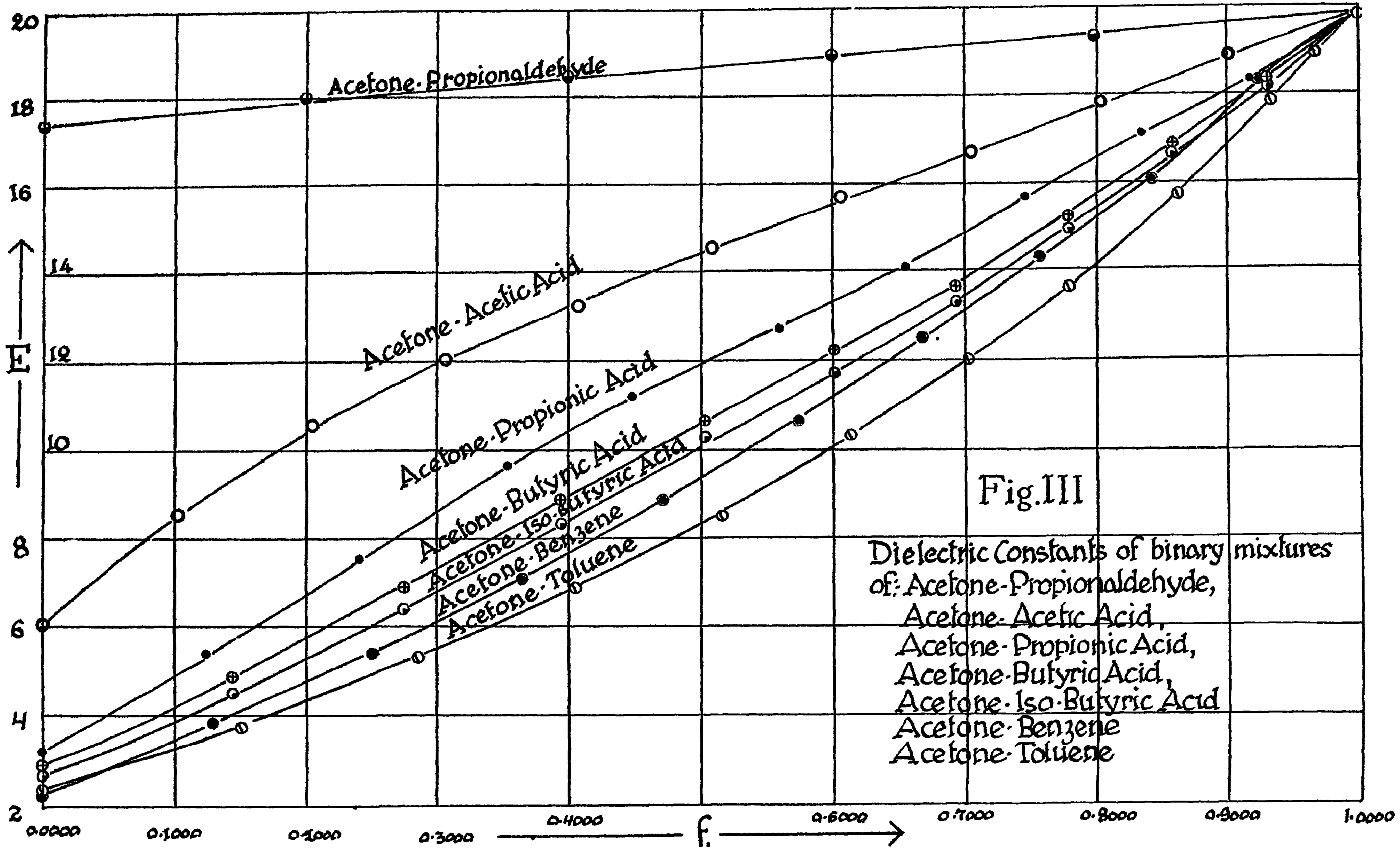


Fig. III

Dielectric Constants of binary mixtures of: Acetone-Propionaldehyde, Acetone-Acetic Acid, Acetone-Propionic Acid, Acetone-Butyric Acid, Acetone-Iso-Butyric Acid, Acetone-Benzene, Acetone-Toluene

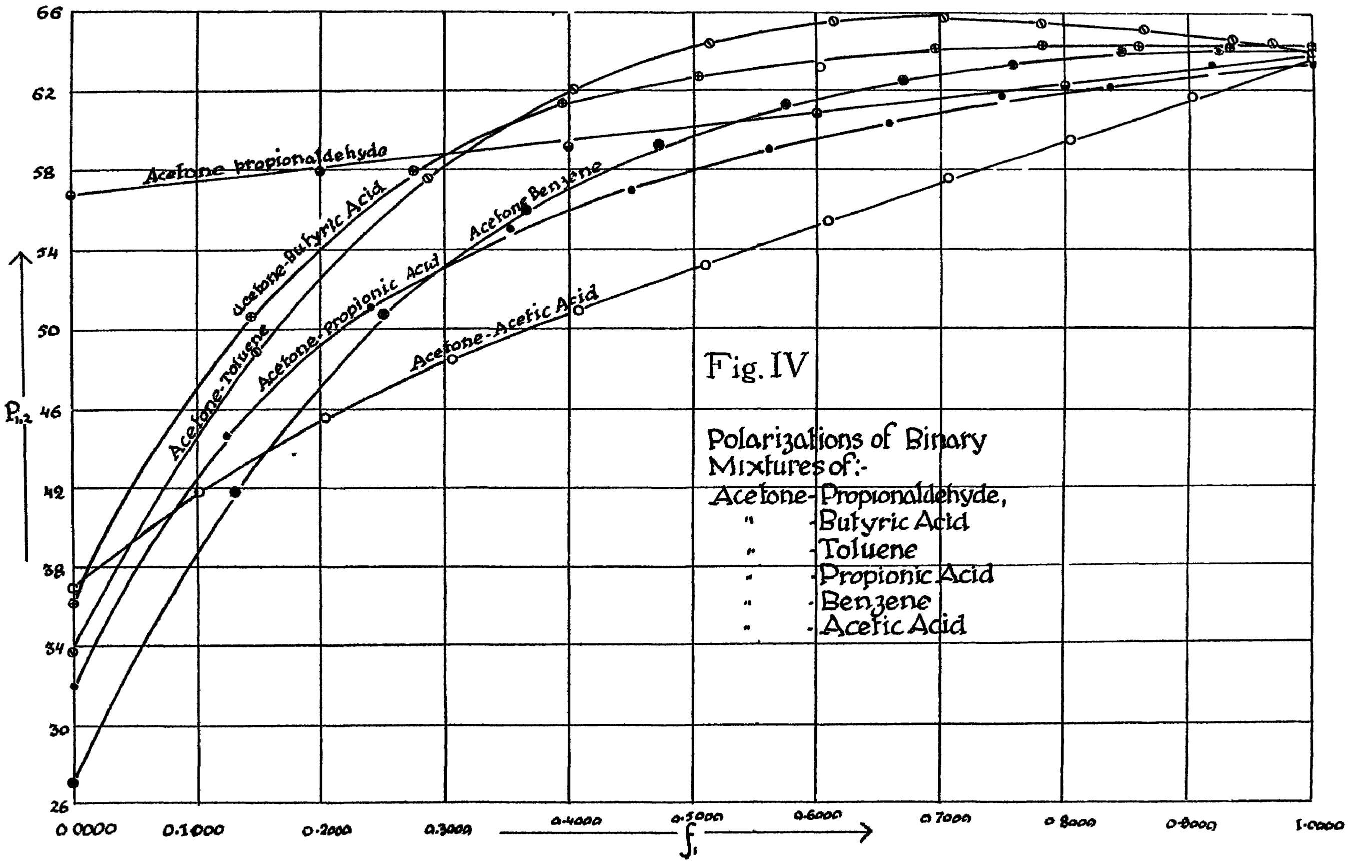


Fig. IV

Polarizations of Binary
Mixtures of:-
Acetone-Propionaldehyde,
" -Butyric Acid
" -Toluene
" -Propionic Acid
" -Benzene
" -Acetic Acid

Graffunder (1931) was convex while the polarization relationship was linear.

The measured dielectric constants of the binary mixtures were compared with the results obtained by calculation using Silberstein's (1885) volume mixture rule for dielectric constants. Deviations from this rule were observed in all cases except in the case of acetone-propionaldehyde mixtures where about 1 per cent discrepancy was observed. The acetone-acetic acid mixtures showed the widest departures from the mixing rules. Kerr observed that this rule did not hold for mixtures of benzene with aniline, benzaldehyde, m. cresol, and benzyl alcohol. In this series of mixtures the second component is in every case associated. Thus a simple explanation is available, since Silberstein's mixture rule formula includes two assumptions: that the capacitances are directly additive and that the molecules of the two components have no dielectric effect on each other. But associated molecules are dipolar and do change even in a non-polar medium.

Marrow (1928) and Stewart (1930) have shown from X-ray diffraction of liquids that a binary mixture does not show the individual diffraction intensities of the components but rather a modification and a single liquid.

When a non-polar material is mixed with a non-polar material, then the mixture polarization $P_{1,\,2}$ is a straight line, and since the molecules do not affect one another, owing to the absence of dipoles, the mixing rule holds. For example, the $P_{1,\,2}$ curve of benzene-toluene has only a slight hump. When a polar and highly associated material like acetone is dissolved in benzene, deviations from the mixing rule and the $P_{1,\,2}$ linearity are to be expected, since, as Debye states, dilution increases the number of free dipoles and hence causes the $P_{1,\,2}$ curve to rise in a convex manner from the composition axis. When acetone was diluted in a slightly polar medium like toluene, a greater curvature of the $P_{1,\,2}$ curve was obtained than in the case of benzene. The acetone-toluene curve rises slightly, with decreasing acetone concentration, to a maximum, because the dipolar effect of acetone and toluene is additive. The acetone-benzene polarization curve has no marked maximum.

Although fatty acids dissolved in non-polar solvents are associated much as they are in the pure state, Broughton (1934) has shown with cryoscopic data that the extent of association of fatty acids decreases slightly with dilution in a non-polar medium like cyclohexane. Acetic acid exists mainly as double molecules in benzene; but with increasing concentration of acetic acid, the polarization increases may be due, following Smyth's reasoning, to the formation of a more highly polymerized form possessed of a moment, the proportion of which increases with increasing concentration of acid.

When two polar-associated materials are mixed, the degree and kind of association present defines the number of dipoles that are free to orient themselves in the field which, together with the moment of the molecules, fix the extent of polarization taking place. Although association is usually regarded as the result of cohesive forces, association arising from dipolar effects may be obtained when molecules lie side by side in such a manner as to tend to neutralize their dipolar effect and produce thereby a quadrapolar unit, or when the molecules align themselves end to end with an additive dipolar effect. Kateswaram observed from the Raman spectrum of acetic acid in the pure state and dissolved both in water and in a non-polar medium like benzene, that aqueous acetic acid shows a clear alteration from the Raman spectrum of pure acetic acid but the Raman spectrum of the non-polar solution remains unaltered. Since increasing the temperature of pure acetic acid produces a Raman spectrum change similar to that produced by dissolving it in water, it is argued that dissolving acetic acid in water causes a change in the association of acetic acid. However, if a change took place in the association of the acid in benzene, it was too small to detect.

When two polar and associated components are being diluted, one in the other, three processes may take place. The association of both materials may remain the same, both may change, or the association of one component may be unaffected while the association of the second changes. Water and acetone are both associated and the mixture polarization relationship is linear. The Raman line of the carbonyl group of acetone in water is not affected and the frequency is the same in the pure liquid as in the water (Edsall, 1936). Fatty acids in water do give a shift which is not due to the carbonyl group of the acid. This Raman frequency shift suggests a possibility of an actual chemical change in the structure of these substances. It appears probable that most of the fatty acids exist largely as double molecules but the presence of two lines of aqueous acetic acid suggests a single and a double molecule of acetic acid. The Raman spectrum of acetic acid in acetone was found by Kateswaram to be similar to the acetic acid-water spectrum. While the dielectric constant curve of acid-acetone (Fig. 3) is distinctly convex, the propionic acid-acetone curve is only slightly convex and both the butyric acid-acetone and isobutyric acid-acetone relationships are slightly concave. The polarization curves, however (Fig. 4), of acetic acid-acetone, propionic acid-acetone, and butyric acid-acetone show the inverse order of convexity where the acetic acid-acetone curve is but slightly convex and the butyric acid-acetone is definitely convex. According to Wilson and Wenzke (1934) the dipole moments do not show any marked increase or decrease in the fatty acid series. While Broughton has shown that in a non-polar medium the extent of association of acetic, propionic, and butyric acids decreases slightly with increasing molecular weight in the fatty acid series, the degree of association which is paralleled by ionization constants does agree with the results obtained in this work, since the $P_{1,\,2}$ of the acetone-fatty acid mixtures were found to decrease in curvature as the molecular weight decreased and as the association and ionization constants increased. The calculated molecular association value of 1.5 for propionaldehyde, using the Chorborski Langinescu (1929) formula, compares favourably with 1.42, the degree of association found by Stakhorski (1929). Although some authors have observed association values for acetone higher than those of propionaldehyde, the Chorborski Langinescu formula used to calculate the propionaldehyde association gives for acetone the value of 1.4. Since propionaldehyde and acetone both have about the same dipole moment, 2.7 and 2.54 × 10^{-18} respectively, and since the degree of association together with the carbonyl properties is similar, a linear dielectric constant and $P_{1,\,2}$ relationship which were found might have been expected.

CONCLUSIONS

In determining dielectric and polarization constants by capacitance measurements, it was observed that conductive materials produce a leak effect on the values of the dielectric constants. Where the leak effect is operative two cases have been recognized:

(*a*) When the capacitance remaining in the standard condenser after the removal of a capacitance equal to the cell is small compared to the cell capacitance, then the error in the measurement of the capacitance will be positive and accordingly the calculated dielectric constant will be too large.

(*b*) When the capacitance remaining in the standard condenser is large compared to the cell, the error in the capacitance will be negative and the dielectric constant will be too small.

The apparatus designed for this work was free from these errors and the extent of the conductance effect on capacitance with this apparatus was observed.

Dielectric constants of acetone, of isobutyric, butyric, propionic, and acetic acids, and of propionaldehyde and toluene were re-determined and found to be in good agreement with those determined by other workers.

Dielectric constants of binary mixtures of acetone-benzene, of acetone-isobutyric, -butyric, -propionic and -acetic acids, and of acetone-propionaldehyde and -toluene were investigated. The polarizations of these mixtures give information comparable to the Raman spectrum results.

REFERENCES

G. Akerlof. J. Amer. Chem. Soc., vol. 54 (1932), 4125.
P. Albright. J. Amer. Chem. Soc., vol. 59 (1937), 2098.
G. Broughton. Trans. Farad. Soc., vol. 30 (1934), 367.
P. Debye. *Dipole moment and chemical structure* (London: Blackie, 1931).
P. Drude. Zeit. physik. Chem., vol. 23 (1897), 267; Ann. Physik, vol. 64 (1898), 131.
J. T. Edsall. J. Chem. Physics, vol. 4 (1936), 1.
W. Graffunder and R. Weber. Zeit. Physik, vol. 65 (1930), 723.
W. Graffunder and E. Heymann. Zeit. Physik, vol. 72 (1931), 744.
M. Grutzmacher. Zeit. Physik, vol. 28 (1924), 324.
P. Kateswaram. Zeit. Physik, vol. 110 (1938), 118.
R. N. Kerr. J. Chem. Soc. (1926), 2796.
G. G. Langinescu. Chem. Rev., vol. 6 (1929), 381.
R. W. Lunt and M. A. G. Rau. Proc. Roy. Soc. (London), vol. A126 (1930), 213.
R. M. Marrow. Phys. Rev., vol. 31 (1928), 10.
A. Piekara and B. Piekara. Compt. Rend , Acad. d. sc., vol. 198 (1934), 1018.
P. Schupp. Zeit. Physik, vol. 75 (1932), 84.
L. Silberstein. Ann. Physik, Neue Folge, vol. 56 (1895), 661.
C. P. Smyth. *Dielectric constants and molecular structure* (New York: Chemical Catalogue Co., 1931).
C. P. Smyth and H. E. Rogers. J. Amer. Chem. Soc., vol. 52 (1930), 1824.
K. M. Stakhorski. Zeit. Elektrochem., vol. 35 (1929), 185.
G. W. Stewart. Phys. Rev., vol. 35 (1930), 726.
C. B. Thwing. Zeit. physik. Chem., vol. 14 (1894), 286.
P. Walden. Zeit. physik. Chem., vol. 46 (1903), 103.
E. W. Washburn. J. Amer. Chem. Soc., vol. 38 (1916), 2436.
J. W. Williams. J. Amer. Chem. Soc., vol. 48 (1926), 1888; vol. 52 (1930), 1831.
C. J. Wilson and H. H. Wenzke. J. Chem. Physics, vol. 2 (1934), 546.

TRANSACTIONS OF

THE ROYAL SOCIETY
OF CANADA

SECTION III

CHEMICAL, MATHEMATICAL,
AND
PHYSICAL SCIENCES

THIRD SERIES—VOLUME XLVI—SECTION III

JUNE, 1952

OTTAWA
THE ROYAL SOCIETY OF CANADA
1952

TABLE OF CONTENTS

Forbidden Transitions in the Spectra of Diatomic Molecules. Presidential Address by G. HERZBERG, F R S C. . . . 1

Dominated Convergence of Kantorovitch Polynomials in the Space L^p. By P. L. BUTZER 23

On the Relation between the P^2*-integral and the Cesàro-Perron Scale of Integrals.* By H W. ELLIS 29

Unbounded Operators and a Theorem of A. Robinson. By G. G. LORENTZ and M. S. MACPHAIL 33

The Extended Euler-Knopp Transformation. By M. S. MACPHAIL 39

Representations of Positive Real Numbers by Infinite Sequences of Integers. By N. S. MENDELSOHN 45

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLVI : SERIES III : JUNE, 1952
SECTION THREE

PRESIDENTIAL ADDRESS

Forbidden Transitions in the Spectra of Diatomic Molecules*

G. HERZBERG, F R S.C.

A. INTRODUCTION

THE study of forbidden transitions in atomic spectra has added a great deal to our understanding of atomic structure, and has contributed greatly to the solution of many astrophysical problems. Similarly the investigation of forbidden transitions in molecular spectra has proved to be of considerable interest, both for a better understanding of molecular structure and for their geophysical and astrophysical applications. Thus one of the most prominent features in the red part of the solar spectrum is now known to be due to a forbidden transition of the O_2 molecule in the earth's atmosphere.

My own interest in this subject dates back more than twenty years when I built my first long absorption tube for the study of forbidden transitions. In this address I should like to review some of the more recent results obtained by my collaborators and myself in the field of forbidden transitions in the spectra of diatomic molecules.

B. ELEMENTARY FACTS ABOUT MOLECULAR SPECTRA

In order to make my discussion intelligible to those not so familiar with the spectra of diatomic molecules, I should like to begin by summarizing very briefly a few elementary facts about the energy levels of diatomic molecules and the selection rules applying to them.

It is well known that the energy of a molecule can be resolved into three parts, the electronic energy E_e, the vibrational energy E_v, and the rotational energy E_r, i.e., we have

$$E = E_e + E_v + E_r = hc\,[T_e + G(v) + F_v(J)]$$

where the factor hc has been taken out in order to convert to term values in cm^{-1} units.

---

*Part of a symposium on Forbidden Transitions.

A molecule, just as an atom, has a series of *electronic states*, but unlike the atom, the electronic states of a diatomic molecule are classified according to the magnitude of the *component of the electronic orbital angular momentum in the direction of the inter-nuclear axis.* According to whether this component, designated by Λ, is 0, 1, 2, . . . (in units $h/2\pi$) the states are called Σ, Π, Δ, . . .states (corresponding to S, P, D, . . . of the atom). Two kinds of Σ states are distinguished, Σ^+ and Σ^-; the former corresponds to the case in which the electronic eigenfunction ψ_e upon reflection at a plane through the internuclear axis remains unchanged, and the latter corresponds to the case where it changes sign.

The *multiplicity* of the electronic states of a diatomic molecule is entirely similar to that of atoms, i.e., if the resultant spin of the electrons is S the multiplicity is $2S + 1$. Thus we have, e.g., a $^3\Delta$ state meaning a state with $\Lambda = 2$ and $S = 1$.

In the case of molecules consisting of two equal atoms (*homonuclear molecules*) like H_2, N_2, O_2, etc., an additional symmetry property must be distinguished. An electronic state of such a molecule is either *even* (indicated by the subscript g) when the electronic eigenfunction remains unchanged upon reflection at the centre, or it is *odd* (indicated by a subscript u) if the eigenfunction changes sign for such a reflection.

In each electronic state the molecule may have a number of *vibrational levels* whose position in a first approximation is given by the simple formula

$$G(v) = \omega_e(v + \tfrac{1}{2}) - \omega_e x_e(v + \tfrac{1}{2})^2$$

where $\omega_e c$ is the (classical) vibrational frequency, $\omega_e x_e \ll \omega_e$ is a small correction term due to the anharmonicity of the vibrations, and v is the vibrational quantum number. According to the formula, the levels are almost equidistant as shown in Fig. 1 (long horizontal lines).

In each vibrational level there is series of *rotational levels* given by the formula

$$F_v(J) = B_v J(J + 1)$$

where the rotational quantum number J assumes the value 0, 1, 2, etc., and represents the total angular momentum in units $h/2\pi$, and where B_v is the rotational constant in the vibrational level v. The rotational levels are not equidistant, but their separation increases with J since $F_v(J)$ is a quadratic function of J. This is also shown in Fig. 1 (short horizontal lines).

The rotational levels are alternatively "*positive*" and "*negative*" depending on whether the total eigenfunction of the molecule remains unchanged or changes sign upon reflection at the origin. In the simplest case of a Σ^+ state the even rotational levels are "positive," the odd rotational levels are "negative." The reverse is the case in a Σ^- state. In Π, Δ, . . . states there are, for each J value, two levels, one "positive" and one "negative."

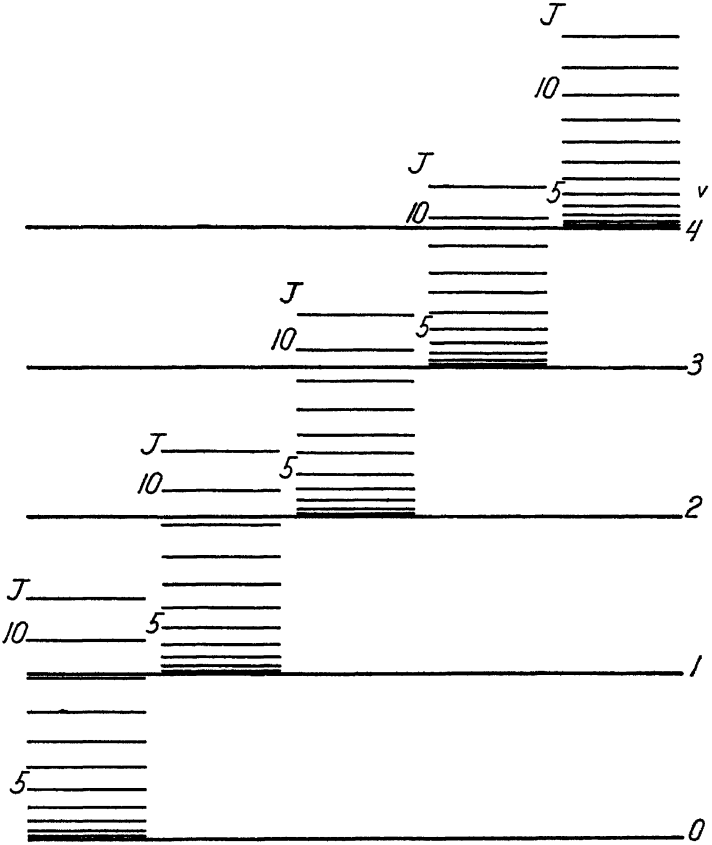


FIGURE 1.—Energy levels of a vibrating and rotating molecule in a given electronic state. The long horizontal lines represent the vibrational levels without rotation ($J = 0$), the short horizontal lines represent the rotational levels for each vibrational level.

In the case of homonuclear molecules we have to distinguish in addition *symmetric* and *antisymmetric* rotational levels corresponding to the behaviour of the eigenfunction with respect to an exchange of the two identical nuclei. For even electronic states (Σ_g, Π_g, . . .) the "positive" rotational levels are symmetric and the "negative" are antisymmetric; for odd electronic states (Σ_u, Π_u, . . .) the reverse is the case.

In the case of doublet, triplet, etc. electronic states one has two, three, etc. series of such rotational levels. The spacing of these doublets and triplets as a function of the rotational quantum number depends on the coupling conditions which we shall not discuss here. However, one special case should be mentioned: if the coupling of the electron spin with the rest of the molecule is very slight the rotational levels are essentially given by a formula as before except that J is replaced by K which is the quantum number of the angular momentum apart from spin.

For the transitions between the various energy levels of a diatomic molecule a number of *selection rules* hold. We may distinguish rigorous selection rules, which hold rigorously for dipole radiation of the free molecule, and approximate selection rules, whose validity depends on the coupling conditions in the molecule.

Rigorous selection rules are the following: the selection rule for the total angular momentum $\Delta J = 0, \pm 1$ (with the restriction that $J = 0$ does not combine with $J = 0$); furthermore, the *parity rule* (corresponding to the Laporte rule for atoms) which says that "positive" rotational levels combine only with "negative" rotational levels (or, for short, $+ \leftrightarrow -$). In the case of homonuclear molecules we have in addition the rule that symmetric rotational levels do not combine with antisymmetric ones (the ortho-para rule, for short, $s \leftarrow\!\!|\!\!\rightarrow a$). Finally, a selection rule may be added which is not independent of those already mentioned, namely, that the dipole moment must be different from zero in order that a pure rotation or a rotation-vibration spectrum can occur.

Examples of *approximate selection rules* are: the rule that states of different multiplicity do not combine with one another (that is, $\Delta S = 0$); the selection rule for the quantum number Λ of the electronic orbital angular momentum $\Delta \Lambda = 0, \pm 1$; and the selection rule for the rotational quantum number K, $\Delta K = 0, \pm 1$. Three other selection rules of this type which are important for the following considerations are the rules that Σ^+ states do not combine with Σ^- states, that even electronic states of molecules with nuclei of equal charge combine only with odd, and that $\Delta K = 0$ does not occur for Σ—Σ transitions.

C. TYPES AND OBSERVATIONS OF FORBIDDEN TRANSITIONS

Two types of forbidden transitions may be distinguished.

1. *Forbidden transitions that violate approximate selection rules.* Examples in atoms are the intercombination lines such as the line λ2537 of mercury.

2. *Forbidden transitions that violate selection rules that are rigorous for dipole radiation in the free molecule.* Such violations may occur for two reasons:

(*a*) In the free molecule the transition may occur as ***magnetic dipole, quadrupole,*** or ***higher pole radiation.*** Well-known examples in the case of atomic spectra are the nebular, auroral, and coronal lines.

(*b*) The transitions may be ***induced by the presence of external fields*** or ***by collisions with other molecules.*** Examples in the case of atomic spectra are the helium lines with $\Delta L = 2$ which occur in strong electric fields.

Forbidden transitions in diatomic molecules of types 1 and 2(*a*) may be observed ***in emission*** under special conditions of excitation, particularly at extremely low pressure as it occurs for example in the upper atmosphere, but also in certain kinds of discharge tubes, either at low pressure, or when an inert gas is added. Forbidden transitions of type 2(*b*) require, of course, the presence of external fields or high pressure to be observed. ***In absorption*** forbidden transitions may be observed by using a very long absorbing path or high pressure or both.

In my laboratory we found the method first suggested by White [1] for the production of a very long absorbing path to be the most satisfactory [2]. In this method by means of three mirrors the light is made to traverse the absorption tube a large number of times. At the Yerkes Observatory I set up an absorption tube of 22 m. length and 25 cm. diameter with which it was possible by means of White's method to obtain up to 250 traversals, i.e., an absorbing path of 5500 m. At the National Research Laboratories in Ottawa we have set up a somewhat smaller tube of 5 m. length in which we have been able to obtain up to 200 traversals corresponding to a path length of 1000 m. In addition, in order to study absorption spectra at high pressures, that is, in order to study forbidden transitions of type 2(*b*) we have built a steel absorption cell of 2 m. length with similar mirrors by means of which we can obtain up to 60 traversals, that is, an absorbing path of 120 m. This tube will stand a pressure of 100 atm. and the whole tube may be immersed in liquid nitrogen in order to study the effect of temperature on these transitions.

D. FORBIDDEN ROTATION-VIBRATION SPECTRA

We shall first consider those forbidden transitions which do not involve a change of electronic state, that is, we shall consider forbidden rotation-vibration spectra of diatomic molecules. In a way the overtone bands of polar molecules, i.e., the transitions from the lowest vibrational level to higher vibrational levels with $v = 2, 3, \ldots$ may be considered as forbidden transitions although this is usually not done. They violate the selection rule for the harmonic oscillator, that is, $\Delta v = \pm 1$. Accordingly at least the higher ones of these transitions with $\Delta v = 3, 4$ can only be obtained with very long absorbing paths. Thus, for example, the 4—0 band of CO requires an absorbing path of 800 m. (see [3]).

Homonuclear molecules have no dipole moment and therefore they do not have an ordinary infrared rotation-vibration specrum. This is also excluded by the selection rules $s \leftarrow\!|\!\rightarrow a$ and $+ \leftrightarrow -$ since without change of electronic state the transitions allowed by the first rule are just the ones rigorously forbidden by the second and conversely. However, a homonuclear molecule does have a *quadrupole moment* which will change during the vibration and may therefore cause the appearance of an extremely weak rotation-vibration spectrum. This is a forbidden transition of the type 2(a) made possible by quadrupole radiation. The selection rules for quadrupole radiation are

$$\Delta J = 0, \pm 1, \pm 2 \text{ and } + \leftarrow\!|\!\rightarrow -.$$

In the case of a transition between the rotational levels of two vi-

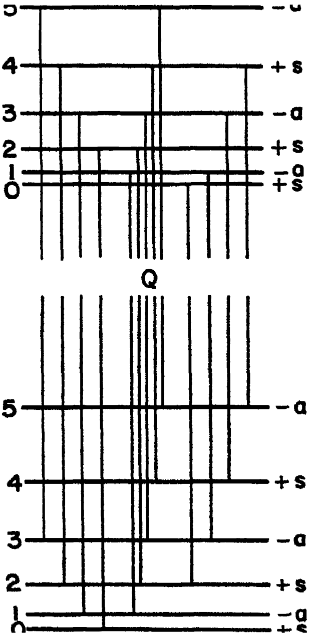


FIGURE 2.—Energy level diagram for the quadrupole rotation-vibration spectrum of a homonuclear molecule.

brational levels of a homonuclear molecule, as shown in Fig. 2, the transitions with $\Delta J = \pm 1$ are eliminated both by the rule $+ \leftarrow\!|\!\rightarrow -$ and the rule that $s \leftarrow\!|\!\rightarrow a$. Therefore, one expects three series of lines or branches, one with $\Delta J = 0$, a Q branch, one with $\Delta J = +2$, an S branch, and one with $\Delta J = -2$, an O branch. In the case of the H_2 molecule, on account of its very small moment of inertia (that is, of the wide spacing of the rotational levels), only very few lines in each of these branches are expected to occur at room temperature, and even fewer lines at lower temperatures. Fig. 3 shows the expected intensity distribution in the 2—0 band for a temperature of 150°K. This temperature was chosen fourteen years ago [4], when this figure was prepared, in order to be able to compare with the spectrum of the major planets in which large amounts of hydrogen were expected. James and Coolidge [5] have made more detailed calculations of the absolute intensities to be expected for these transitions and on the basis of their calculations one finds that the 1—0, 2—0, and 3—0 bands would require 2.5, 2.7, and 13.0 km. path at atmospheric pressure respectively.

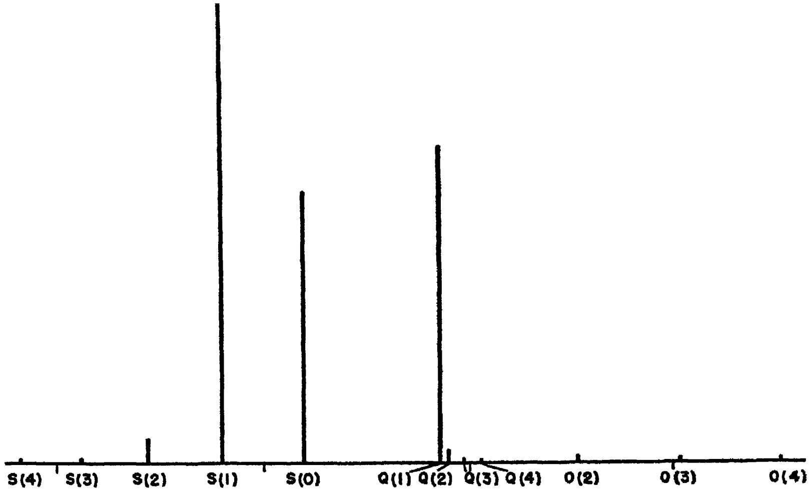


Figure 3.—Predicted structure and intensity distribution of the 2—0 quadrupole band of H_2 at 150°K. The height of the vertical lines indicates the predicted intensity.

With the absorption tube at Yerkes Observatory, which I mentioned previously, it appeared possible to observe the *quadrupole spectrum of hydrogen* since by increasing the pressure to 10 atm. an equivalent path of up to 55 km. atm. could be obtained. When this was done the above prediction was indeed verified in that the 2—0 band was observed with a path length of 10 km. atm. and the 3—0 band with a path length of 40 km. atm. [6]. Fig. 4 shows several of the observed lines. Unfortunately, the spectral region in which these lines occur is overlapped by absorption due to water vapour and therefore in each case a comparison spectrum without hydrogen is shown. In this way the quadrupole lines of hydrogen can easily be identified. It should be particularly noted that in spite of a pressure of about 10 atm. the quadrupole lines are extremely sharp. Their width was as small as could be obtained in the second order of a 21 ft. grating.

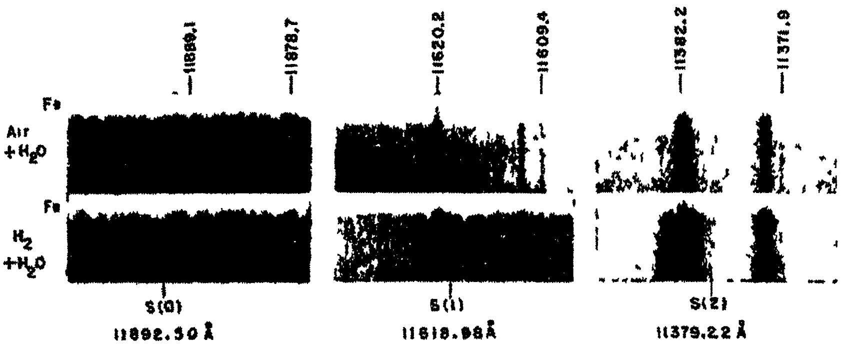


FIGURE 4.—Lines of the 2—0 quadrupole band of H_2. The top strips show for comparison the spectrum of moist air. In this way the quadrupole lines of H_2 in the bottom strips can be identified in spite of the overlapping H_2O spectrum

A slight systematic shift was observed between the predicted positions of the quadrupole lines (predicted from the rotational and vibrational constants of hydrogen derived from its ultraviolet spectrum). This shift is due to the low accuracy obtainable in the far ultraviolet. By using the quadrupole spectrum it was possible to obtain greatly improved rotational and vibrational constants of the hydrogen molecule in its ground state. These data are of particular interest from the point of view of a calculation of the thermodynamic properties of hydrogen.

Unless the amounts of hydrogen present in the atmospheres of the major planets are very much larger than those used in the laboratory, spectra of very high resolution will be required in order to detect hydrogen by means of these quadrupole lines in planetary spectra. Such high resolution has not yet been accomplished. But Kuiper [7], at Yerkes Observatory, has observed a diffuse feature in the spectra of Uranus and Neptune at 8267Å which is exactly at the position of the line *S*(0) of the 3—0 quadrupole band. This line would be the strongest line expected at very low temperature. However, the considerable width of the planetary feature (about 40Å) makes it impossible to identify it as the quadrupole line.

More recently the rotation-vibration spectra of hydrogen as well as of other homonuclear molecules have been observed by Crawford, Welsh, Locke, and their collaborators [8; 9] at Toronto in absorption *at high pressure*, i.e., as forbidden transitions of type 2(*b*). These spectra are very different from the quadrupole spectrum in that their intensity increases with the square of the pressure and that in place of sharp lines very broad features (about 300 cm^{-1} wide) are observed.

Condon [10], in 1932, predicted a rotation-vibration spectrum of homonuclear molecules induced by a constant electric field. This spectrum can also be considered as a Raman spectrum at zero exciting frequency. The same three branches are expected as occur in the quadrupole spectrum. It appears that the spectra observed by Crawford and Welsh correspond to this spectrum except that the external electric field is replaced by the collisions with other molecules.

Crawford and Welsh observed that the *line width* of this pressure-induced spectrum decreases with decreasing temperature. This is shown for the 2—0 band in Fig. 5. While for the fundamental (1—0 band) at low temperature essentially three lines occur, $Q(1)$, $S(0)$,

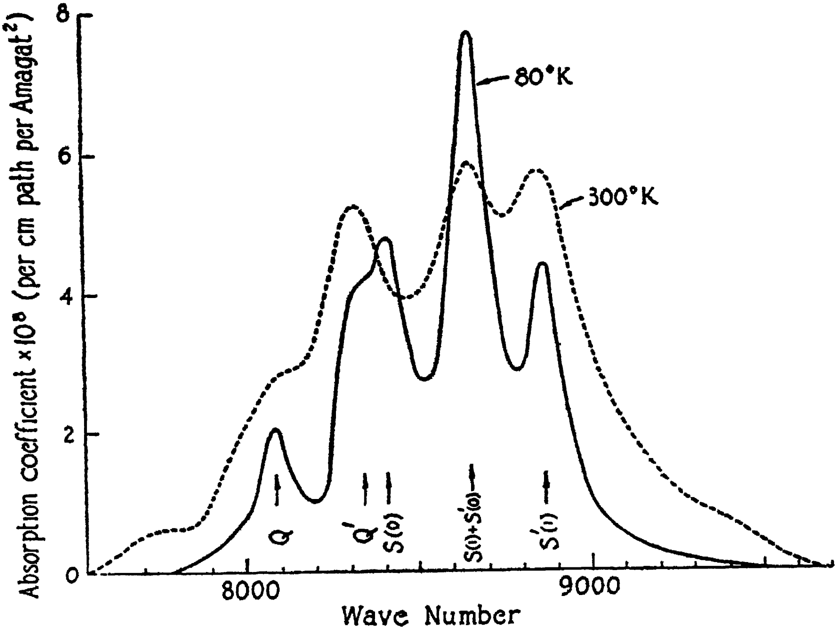


Figure 5.—Induced infrared absorption of H_2 near 8500 cm^{-1} at 300° and 80°K according to Welsh, Crawford, MacDonald, and Chisholm [9].

and $S(1)$, for the overtone (2—0 band) some additional lines, marked Q', $S'(0)$, $S'(1)$ appear (see below).

Since the positions of the pressure-induced lines of the 1—0 and 2—0 bands agree closely with those of the quadrupole lines it appeared probable [11] that the planetary feature observed by Kuiper at 8267Å whose position agrees with $S(0)$ of the 3—0 quadrupole band might be the $S(0)$ line of the pressure-induced 3—0 band. A test of this assumption was therefore made with the high pressure, low temperature absorption tube mentioned in the introduction [12]. In agreement with the results of Welsh and Crawford, at room temperature the 3—0 band was found to consist of very diffuse features. When, however, the temperature was lowered to that of liquid nitrogen a much clearer and more definite spectrum was obtained which is shown in Fig. 6*a*. In this spectrogram actually more lines are seen than

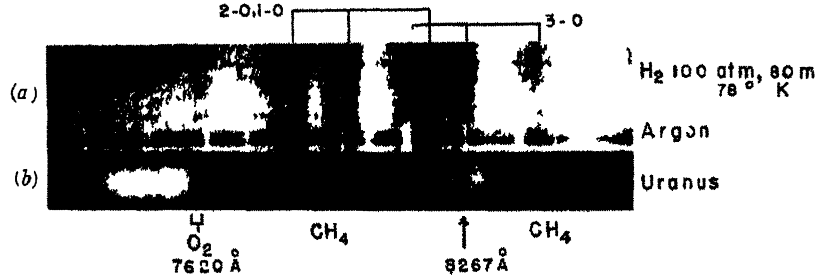


FIGURE 6.—Induced infrared absorption of H_2 near 8270Å observed (*a*) in the laboratory and (*b*) in the spectrum of Uranus after Kuiper. The three diffuse lines marked 3—0 are $Q(1)$, $S(0)$, and $S(1)$ (from right to left) of the single transition 3—0. The three lines marked 2—0, 1—0 are $Q(1)$, $S(0)$, and $S(1)$ of the double transition.

expected. The expected lines are the three lines at the right, i.e., $Q(1)$, $S(0)$, and $S(1)$. The two lines at shorter wavelengths cannot be accounted for as lines of the ordinary pressure-induced spectrum. Similar additional lines occur in the 2—0 band (Fig. 5) although they are more overlapped by the normal lines. They have been explained by Welsh and Crawford as due to a *double transition*, that is, a transition in which each of the two colliding molecules goes from the lowest vibrational level to that with $v = 1$. In other words, the wave number of this transition is twice the wave number of the 1—0 band with rotational energy added in the same way as for the single (2—0) transition. Similarly the extra lines in the region of the 3—0 band (which are actually stronger than the lines of the single transition) can be accounted for as due to a double transition in which one of the collision partners goes from $v = 0$ to 1, the other simultaneously from $v = 0$ to 2. Table I shows the wavelengths predicted on this basis as well as those of the observed lines. The central line has to be assigned to both the single band and the double band. The agreement between observed and calculated lines is so good that hardly a doubt can remain as to the correctness of the interpretation of the spectrum as due to a single and a double transition even though it appears rather difficult to account on a theoretical basis for the occurrence of such a double transition.

TABLE I

OBSERVED AND PREDICTED PRESSURE-INDUCED SPECTRUM OF H_2

Designation		Observed λ (Å)	Predicted λ (Å)
Single transition (3—0)	$Q(0)$	8469	8488
	$Q(1)$		8497
	$S(0)$	8258	8273
	$S(1)$	8166	8151
Double transition (2—0, 1—0)	$Q(0)$		8165
	$Q(1)$		8171
	$S(0)$	7958	7956
	$S(1)$	7830	7832

It appears to be significant that the line $S(0)$ of the single transition agrees closely ***both in wavelength and in width*** with the feature 8267Å observed in the spectra of Uranus and Neptune. Fig. 6*b* shows, for comparison with Fig. 6*a*, one of Kuiper's Uranus spectra on which the 8267Å feature can be seen. With one exception the other lines of the H_2 spectrum are obscured in the planetary spectra by strong CH_4 bands. The line 8166Å is not completely obscured by CH_4 and there is indeed evidence that it is present in the Uranus spectrum. However, it appears to be considerably weaker relative to 8267Å than in the laboratory spectrum. This difference can be accounted for by the assumption of the presence of large amounts of helium in addition to hydrogen, since collisions of He and H_2 would enhance the single transition but not the double transition. At any rate it appears that the presence of large amounts of H_2 in the atmospheres of Uranus and Neptune is now established by means of the pressure-induced H_2 spectrum. The amount of H_2 necessary to account for the observed intensity of the 8267Å feature in Uranus can be roughly obtained by comparison with the laboratory data on the basis of the quadratic dependence on the pressure. Assuming a temperature of 78°K, a helium-hydrogen ratio of 3 : 1, and therefore a scale height of 18 km., one finds a partial pressure of 2 atm. of H_2 at the bottom of the visible atmosphere.

Another example of a forbidden rotation-vibration spectrum is the *rotation-vibration spectrum of* HD In its equilibrium position the HD molecule, just as the H_2 molecule, has no dipole moment and therefore no rotation-vibration spectrum would be expected. This spectrum is also forbidden by the selection rule that even electronic states combine only with odd ones, a rule that applies to all molecules whose nuclei have equal charge. However, when the nuclei are vibrating, as shown in Fig. 7, since the centre of mass remains at rest but is no longer symmetrically placed, both the centre of the "positive" and that of the "negative" charges move back and forth. On account of the inertia of the electrons the centre of the "negative" charges will lag behind that of the "positive" charges, and therefore an oscillating dipole moment arises which in turn produces a rotation-vibration spectrum. Wick [13] about fifteen years ago calculated the intensity of the 1—0 band of this spectrum According to his formulae it

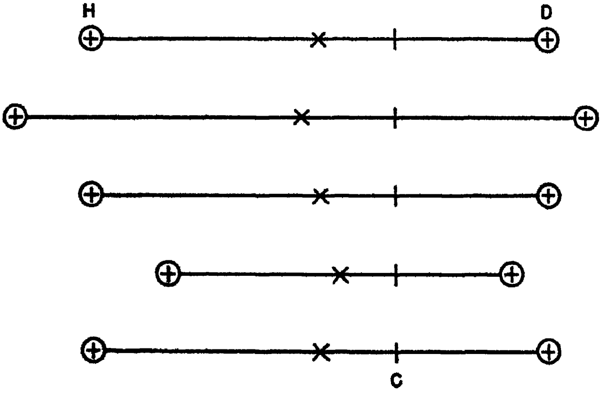


FIGURE 7.—Successive stages of vibration of HD showing origin of dipole moment. The centre of mass is at *C*, the centre of both positive and negative charges would be at X if the inertia of the electrons were neglected.

would require an absorbing path of the order of 30 m. atm. to detect this band. With the absorption tube described earlier it has been possible to observe the 3—0 and 4—0 bands of HD at 9650 and 7420Å using an absorbing path of 1000 m. at a pressure of 1 atm [14]. Fig 8 shows the 4—0 band. Unlike the quadrupole spectrum of hydrogen, in the present case one has the usual simple structure of a *P* and an *R* branch which are clearly visible. By an exact measurement of these bands it has been possible to obtain improved molecular constants of the HD molecule.

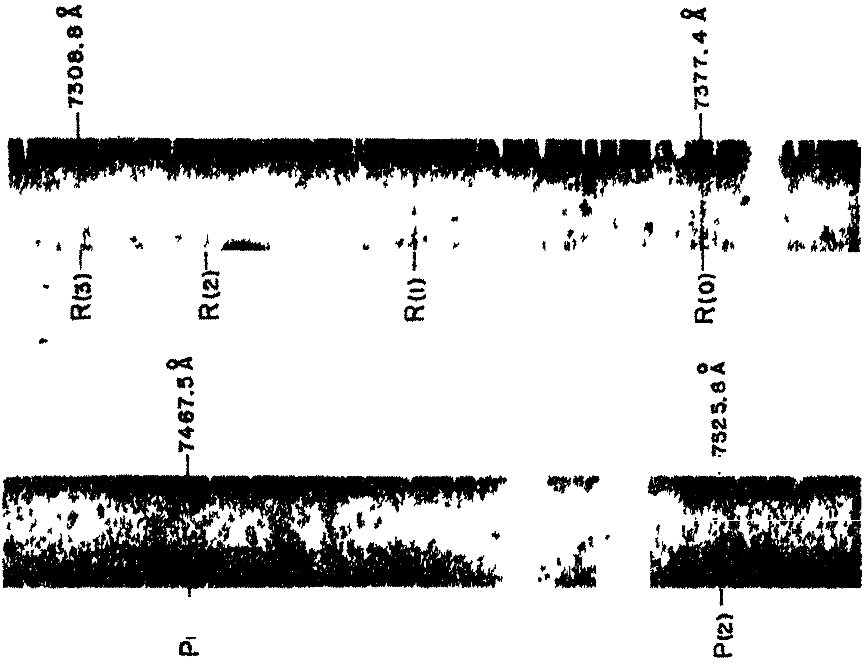


FIGURE 8 — 4—0 rotation-vibration band of HD.

A surprising result is the comparatively small difference in intensity of the 3—0 and 4—0 bands of HD contrary to the usual very rapid decrease of intensity in the series of overtones. This small intensity decrease in the case of HD is apparently connected with the fact that the dipole moment is produced by the vibration. Quantum-theoretical calculations by Wu [15] at Ottawa have indeed shown quantitatively that such a small intensity decrease should be expected. It will be interesting to observe the pressure-induced spectrum of HD, but such experiments have not yet been carried out.

E. Forbidden Electronic Band Spectra

Let us turn now to the discussion of forbidden electronic band spectra, that is, transitions in which the molecule goes from one electronic state to another such that one or more of the electronic selection rules are violated. The most common type of forbidden electronic transition in diatomic molecules is the *intercombination between two states of different multiplicity* (that is, a transition violating the approximate selection rule $\Delta S = 0$). The first case of this type that was observed and identified as such was the *Cameron system of CO* which represents a $^3\Pi - {}^1\Sigma$ transition, the lower state being the ground state of the molecule [16]. These bands have been observed both in emission and in absorption. In absorption they require an absorbing path of about 1 m. at atmospheric pressure. Fig. 9 shows

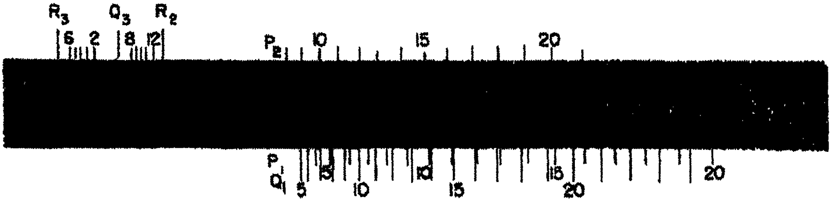


FIGURE 9.— 0—0 Cameron band of CO in absorption.

the structure of one of the bands as obtained in absorption [17]. Each band consists of three overlapping components of three branches each, corresponding to the three components of the $^3\Pi$ state.

Transitions similar to the Cameron bands of CO have been observed for many of the halides of the elements of the third column of the periodic system by Miescher and Wehrli [18]. At Ottawa recently, Sharma [19] observed the corresponding transitions in AlCl and AlBr. In these cases the triplet splitting in the $^3\Pi$ state is very large compared to the rotational constant B, and therefore a coupling case applies in which only two of the three components of the $^3\Pi$ state combine with the ground state. These are the transitions $^3\Pi_1 - {}^1\Sigma$ and $^3\Pi_0 - {}^1\Sigma$ of which the former gives rise to bands with P, Q, and R branches, whereas the latter gives bands with P and R branches only.

Another important case is the so-called *Vegard-Kaplan bands of nitrogen* which represent a $^3\Sigma_u^+ - {}^1\Sigma_g^+$ transition, the lower state being the ground state of the N_2 molecule. These bands have been observed in emission in certain electric discharges [20; 21; 22; 23] as well as in the spectrum of the night sky [24; 25]. One of the first objects of my own work on absorption spectra with long absorbing paths was the observation of the Vegard-Kaplan bands of nitrogen in absorption in order to verify their assignment as a ${}^3\Sigma_u{}^+ - {}^1\Sigma_g{}^+$ transition and also in order to determine their absolute intensity. These absorption bands are expected to be in the vacuum region and it is not easy to be sufficiently free from the much stronger absorption due to oxygen in order to observe them. Recently with our new vacuum spectrograph at Ottawa we have tried again to observe these bands in absorption, and, using very pure nitrogen, have been able to ascertain that with an absorbing path of 6 m. at atmospheric pressure they do not occur. Apparently the transition probability for this transition is very much lower than that of the Cameron bands of CO. The difference between the two cases seems to lie in the fact [26] that in the case of CO there is a state, ${}^1\Pi$, of the same electron configuration which combines strongly with the ground state and from which the forbidden ${}^3\Pi - {}^1\Sigma$ transition can borrow intensity, while in the case of N_2 the excited electron configuration does not give rise to a state that combines strongly with the ground state. Another example of this type is an extremely weak absorption system of oxygen only recently discovered among the forbidden $\Sigma^+ - \Sigma^-$ bands to be discussed later. This transition is of the type ${}^1\Sigma_u{}^- - {}^3\Sigma_g{}^-$.

We have tried to find intercombinations in a number of other diatomic molecules: NO, for which a ${}^4\Pi - {}^2\Pi$ is expected, SiO, for which the analogue of the Cameron bands of CO, ${}^3\Pi - {}^1\Sigma$, is expected; NH, for which a ${}^1\Sigma - {}^3\Pi$ transition is expected; and CH, for which various intercombinations involving the low lying but unobserved ${}^4\Sigma^-$ state are expected. In all these cases our search has thus far been unsuccessful. Further search with improved means would seem important in order to round out our knowledge of these molecules.

While the intercombinations are forbidden transitions of type 1, there are also a number of known cases of forbidden transitions of type 2(*a*), that is, transitions rigorously forbidden as electric dipole radiation but actually occurring weakly on account of the possibility of *magnetic dipole radiation*. The best known case of this type is the *red atmospheric oxygen bands* which, as was shown by Mulliken [27] and Van Vleck [28], represent a ${}^1\Sigma_g{}^+ - {}^3\Sigma_g{}^-$ transition. The structure of the bands is explained in Fig. 10. Since the oxygen nucleus has zero spin only the symmetric rotational levels occur which are, in the upper state, the levels with even K, and in the lower state, those with odd K. On account of the selection rule $\Delta J = 0, \pm 1$, each one of the upper state rotational levels can combine with only two of the three triplet components of a given lower triplet, thus giving rise to two series of apparent doublets.

Another example whose structure is also explained in Fig. 10 is the *infrared atmospheric oxygen bands* which represent a $^1\Delta_g$ — $^3\Sigma_g^-$ transition. In this case rotational levels occur for each J value of the upper state $^1\Delta_g$ and, therefore, additional branches occur, one each in which the rotational quantum number K changes by ± 2, and three branches in which it changes by 0. All these branches have been observed [29] in the 0—0 band which occurs prominently in the solar spectrum at 12700Å. While the red atmospheric oxygen bands can be obtained in absorption with a path length of 6 m. at atmospheric pressure, the infrared bands require a much longer path of the order of 50 m. The reason for this difference is clear when it is considered that the infrared bands, in addition to being a magnetic dipole intercombination transition, violate the rule $\Delta\Lambda = 0, \pm 1$. The red atmospheric oxygen bands have also been observed in emission, both in the spectrum of the night sky [30] and in the laboratory in certain flames and afterglows [31; 32].

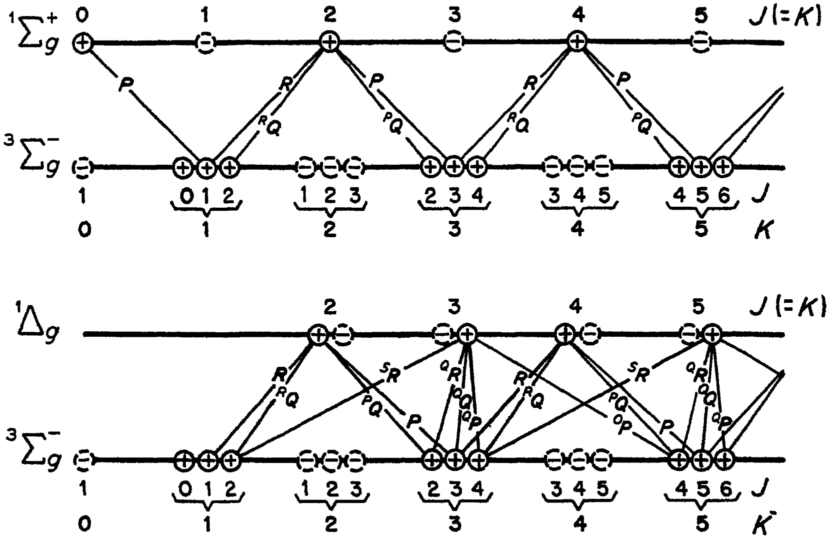


FIGURE 10.—Origin of branches in $^1\Sigma_g^+$ — $^3\Sigma_g^-$ and $^1\Delta_g$ — $^3\Sigma_g^-$ bands.

Still another example of a magnetic dipole transition is the system of far ultraviolet absorption bands of the nitrogen molecule usually referred to as *Lyman-Birge-Hopfield bands* [33]. They extend from 1450Å to shorter wavelengths and represent the first moderately strong absorption system of the N_2 molecule. In order to observe this system an absorbing path of about 40 cm. at atmospheric pressure is required.

It was recognized only a few years ago [34] that these bands represent a magnetic dipole transition of the type $^1\Pi_g - {}^1\Sigma_g^-$. At Ottawa we have recently re-investigated the Lyman-Birge-Hopfield bands [35] in emission in an electric discharge at low pressure and have established the presence of a predissociation in the upper level, at an energy of about 9.7 e.v. This predissociation is of some importance because it will lead to the production of free nitrogen atoms in the upper atmosphere by solar radiation below 1270Å.

The first long absorption tube that I built over twenty years ago gave an absorbing path of 25 m. When it was filled with oxygen at atmospheric pressure, I found a new very weak absorption band system of oxygen in the region 2500Å [36]. The bands of this system appear under medium dispersion to consist only of Q branches and were therefore ascribed to a $\Sigma^+ - \Sigma^-$ transition, i.e., a forbidden transition of type 1. In the present case the approximation to which the selection rule ($\Sigma^+ \leftarrow\!|\!\rightarrow \Sigma^-$) holds is extremely good, and these bands are even weaker than magnetic dipole transitions. More recently [37] this band system has been re-investigated under high dispersion with the much longer absorbing paths set up at Yerkes Observatory. More bands at longer wavelengths have been found and the fine structure has been completely resolved. Fig. 11 shows part of one of the bands under high resolution. It is seen that each of the "lines" of the Q branch is resolved into six component lines and, in addition, some fainter lines belonging to different branches are present. The fine structure is in

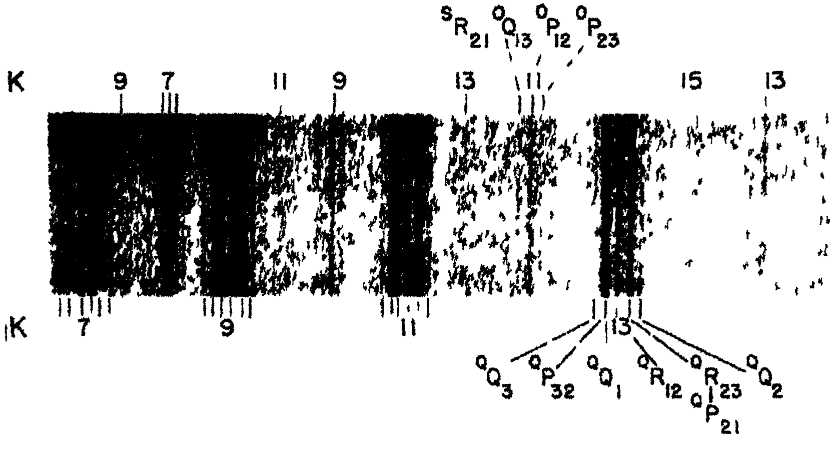


FIGURE 11.—Part of the 9—0 band of the $^3\Sigma_u^+ - {}^3\Sigma_g^-$ system of O_2 under high resolution. Under low dispersion the groups of lines marked K = 7, 9, 11, 13 below form one apparently single Q branch.

agreement with the assumption that this transition is a ${}^3\Sigma_u{}^+ -- {}^3\Sigma_g{}^-$ transition which, although forbidden by the selection rule $\Sigma^+ \leftarrow\mid\rightarrow \Sigma^-$, occurs weakly on account of spin-orbit interaction. Fig. 12 gives an energy level diagram for such a transition. It is found that the triplet splitting in the upper state is about twice as large as in the ground state of O_2 and has the opposite sign. Applying the selection rule $\Delta J = 0, \pm 1$, it is readily seen from this diagram how seven Q branches arise, two of which coincide, and how, in addition, branches with $\Delta K = \pm 2$ are to be expected, i.e., O and S form branches. The observed relative intensities of the various branches do not agree well with those predicted from theory. However, on account of the closeness to the dissociation limit, special coupling conditions apply which were not considered in the theoretical prediction. It may be of interest to recall that this band system has been identified as occurring in emission in the light of the night sky [38; 25].

The ${}^3\Sigma_u{}^+ — {}^3\Sigma_g{}^-$ bands observed in absorption form a simple progression with a clear convergence limit at 2425Å. The energy of this dissociation limit is slightly higher than expected from the well-known dissociation limit in the Schumann-Runge bands of oxygen by subtracting the excitation energy of the ${}^1D$ state of atomic oxygen. The reason for this discrepancy may lie in the fact that the dissociation products of the ${}^3\Sigma_u{}^+$ state are not ${}^3P_2 + {}^3P_2$, but ${}^3P_2 + {}^3P_{1 \text{ or } 0}$. However, if that were the case, a contradiction to the non-crossing rule would arise and it is perhaps more likely that the fault lies with the slight extrapolation made in obtaining the dissociation limit of the Schumann-Runge bands. If this were correct the dissociation energy of the O_2 molecule would have to be changed from the value 5.080 e.v. now usually quoted to 5.109 e.v.

Finally, I should like to consider examples of forbidden electronic transitions of type 2(*b*) which have also been observed for the oxygen molecule. If the absorption spectrum of oxygen is studied *at higher pressures*, a number of completely diffuse bands are observed in the visible region in addition to the ordinary atmospheric oxygen bands. The strongest of these bands occur at 6300, 5780, 4775, 4475, 3815, and 3617Å. Even under high resolution there is no indication of any fine structure. These bands can be observed at comparatively low pressures if the path is sufficiently long. The intensity of these bands increases with the square of the pressure and they have generally been ascribed to the presence in oxygen of O_4 molecules. As was first shown by Ellis and Kneser [39], all these bands can be well accounted for, as far as their wavelengths are concerned, as due to *double electronic transitions* in an O_4 molecule or quasi molecule. If in both O_2 molecules forming

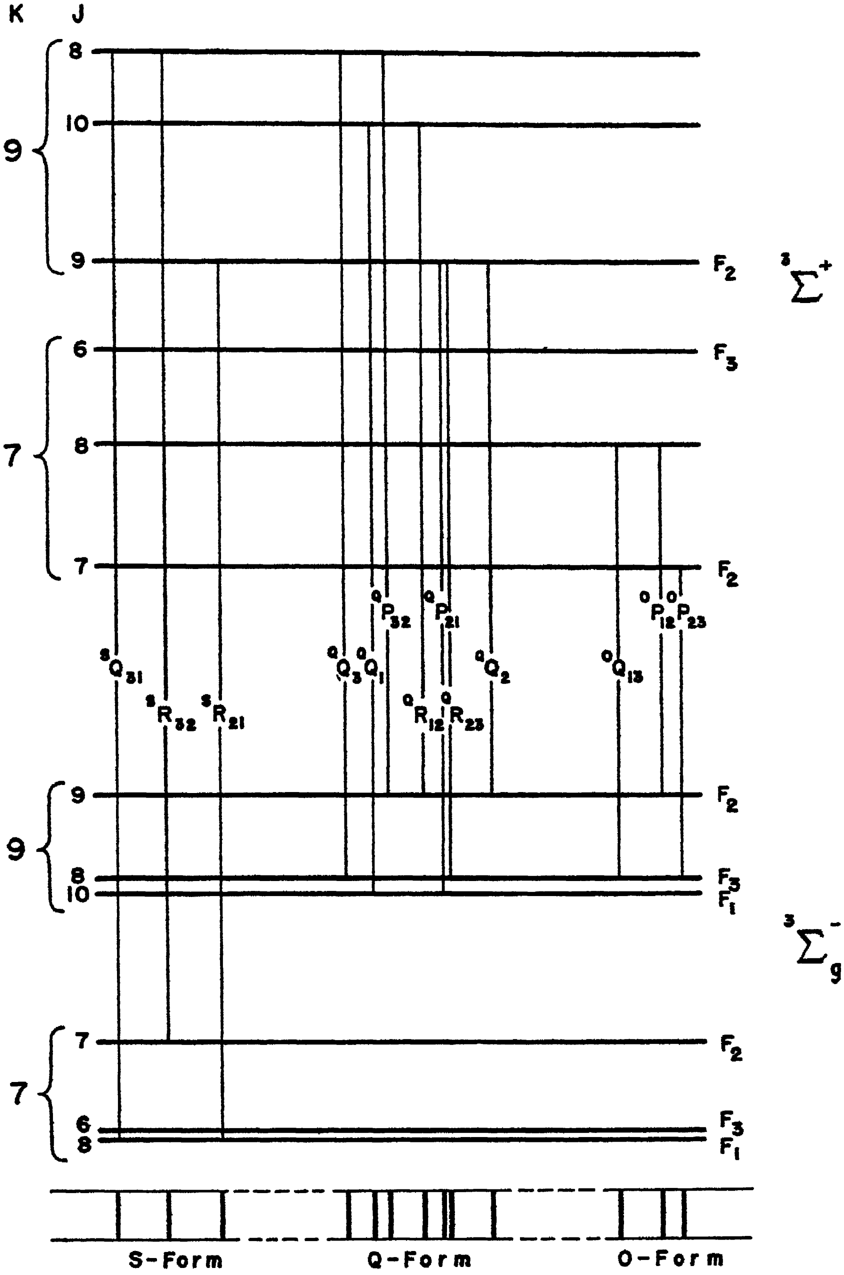


FIGURE 12.—Energy level diagram for a ${}^3\Sigma_u^+ - {}^3\Sigma_g^-$ band. The strip at the bottom gives schematically the resulting spectrum. Only one line of each branch is shown.

an O_4 molecule a transition to the $^1\Delta$ state takes place simultaneously, the resulting wavelength is 6342Å which agrees with one of the strong diffuse bands. Similarly, if in one of the O_2 molecules a transition to the $^1\Delta$ state takes place, and in the other a transition to the $^1\Sigma$ state, the resulting wavelength is 4760Å; and if in both O_2 molecules the transition to $^1\Sigma$ takes place the wavelength is 3810Å. Both these wavelengths again agree with two of the observed diffuse bands. The remaining bands can be accounted for by adding vibrational quanta of the $^1\Delta$ and $^1\Sigma$ state. It would appear that there is a close analogy of these double transitions of oxygen to the double transitions in the rotation-vibration spectrum of hydrogen discussed previously.

In addition to the diffuse bands in the visible region, there occurs at high pressure a progression of ultraviolet bands which almost coincide with the $\Sigma^+ - \Sigma^-$ bands of oxygen [40; 41]. However, these pressure-induced bands may be observed simultaneously with the $\Sigma^+ - \Sigma^-$ bands and cannot, therefore, be just pressure-broadened $\Sigma^+ - \Sigma^-$ bands. Fig. 13 shows a spectrum in which both systems can be seen.

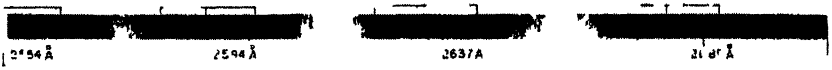


FIGURE 13.—Near ultraviolet absorption of O_2 at 8 atm. pressure, and a path length of 90 m. The discrete $^3\Sigma_u^+ - ^3\Sigma_g^-$ bands are marked at the bottom, the diffuse triplet bands of Wulf and Finkelnburg and Steiner are marked at the top.

Each of the pressure-induced bands has three maxima, but otherwise is completely diffuse. With the long paths used in the more recent work it has been possible to observe at lower pressures a system of very faint and rather complicated bands which may be the exact analogue in free O_2 of the pressure-induced bands. These bands have tentatively been assigned to a $^3\Delta_u - ^3\Sigma_g^-$ transition in oxygen.

F. CONCLUSION

Just as for diatomic molecules, forbidden transitions are of considerable importance for the study of polyatomic molecules. Singlet-triplet intercombinations, as well as other forbidden transitions of type 1, have been found for them (for example, the well-known near-ultraviolet absorption bands of C_6H_6). As far as I am aware, no forbidden transitions of type 2(*a*), i.e., magnetic dipole and quadrupole transitions, are known for polyatomic molecules. Pressure-induced infrared bands of CO_2, i.e., forbidden transitions of type 2(*b*), have been observed by Welsh and Crawford [8]. Both for polyatomic and for diatomic molecules a good deal of work remains to be done on forbidden transitions. Experimentally, the discovery of new forbidden transitions and the study of their intensities and of the conditions of their occurrence will contribute to a better understanding of the structure of the molecules considered and to the solution of certain astrophysical and chemical problems. Further theoretical work on forbidden transitions is needed in order to understand such phenomena as the occurrence of double transitions in collisions between two hydrogen molecules and two oxygen molecules, the intensity distribution in the branches of forbidden bands, and the absolute intensity of forbidden transitions.

REFERENCES

1. J. U. White, J. Opt. Soc., vol. 32 (1942), 285.
2. G. Herzberg, in G. P. Kuiper's *The Atmospheres of the Earth and Planets* (Chicago, 1949), p. 346.
3. G. Herzberg and K. N. Rao, J Chem. Phys., vol. 17 (1949), 1099.
4. G. Herzberg, Astrophys. J., vol. 87 (1938), 428.
5. H. M. James and A. S. Coolidge, Astrophys. J., vol. 87 (1938), 438.
6. G. Herzberg, Nature, vol. 163 (1949), 170; Can. J. Res., vol. A28 (1950), 144.
7. G. P. Kuiper, Astrophys. J., vol. 109 (1949), 540; *The Atmospheres of the Earth and Planets* (Chicago, 1949), p. 325; second edition (1952), p. 384.
8. H. L. Welsh, M. F. Crawford, and J. L. Locke, Phys. Rev., vol. 76 (1949), 580; M. F. Crawford, H. L Welsh, J. C. F. MacDonald, and J. L. Locke, Phys. Rev., vol. 80 (1950), 469.
9. H. L. Welsh, M. F. Crawford, J. C. F. MacDonald, and D. A. Chisholm, Phys. Rev., vol. 83 (1951), 1264.
10. E. U. Condon, Phys. Rev., vol. 41 (1932), 759.
11. G. Herzberg, J. Roy. Astron. Soc. Can., vol. 45 (1951), 100.
12. G. Herzberg, Astrophys.J., vol. 115 (1952), 337.
13. G. C. Wick, Atti Reale Accad. Lincei, vol. 21 (1935), 708.
14. G. Herzberg, Nature, vol. 166 (1950), 563.
15. T. Y. Wu, Can. J. Phys., vol. 30 (1952), 291.
16. J. J. Hopfield and R. T. Birge, Phys. Rev., vol. 29 (1927), 922.
17. L. Gerö, G. Herzberg and R. Schmid, Phys. Rev., vol 52 (1937), 467.
18. E. Miescher and M. Wehrli, Helv. Phys. Acta, vol. 7 (1934), 331, 298.
19. D. Sharma, Astrophys. J., vol. 113 (1951), 210, 219.
20. L. Vegard, Z. Physik, vol. 75 (1932), 30.
21. J. Kaplan, Phys. Rev., vol. 44 (1933), 947; vol. 45 (1934), 675.
22. O. R. Wulf and E. H. Melvin, Phys. Rev., vol. 55 (1939), 687.
23. J. Janin, Ann. de Phys. (12), vol. 1 (1946), 538.
24. J. Cabannes and J. Dufay, C. R., Paris, vol. 200 (1935), 1504.
25. P. Swings and A. B. Meinel, in G. P. Kuiper's *The Atmospheres of the Earth and Planets*, second edition (Chicago, 1952), p. 159.
26. R. S. Mulliken, private communication.
27. R. S. Mulliken, Phys. Rev., vol. 32 (1928), 880; Rev. Mod. Phys., vol. 4 (1932), 1.

28. J. H. Van Vleck, Astrophys. J., vol. 80 (1934), 161.
29. L. Herzberg and G. Herzberg, Astrophys. J., vol. 105 (1947), 353.
30. A. B. Meinel, Publ. Astron. Soc. Pac., vol. 60 (1948), 373; Trans. Amer. Geophys. Un., vol. 31 (1950), 21; Astrophys. J., vol. 112 (1950), 464.
31. J. Kaplan, Phys. Rev., vol. 71 (1947), 274.
32. R. C. Herman, H. S. Hopfield, G. A. Hornbeck, and S. Silverman, J. Chem. Phys., vol. 17 (1949), 220.
33. R. T. Birge and J. J. Hopfield, Astrophys. J., vol. 68 (1928), 257.
34. G. Herzberg, Phys. Rev., vol. 69 (1946), 362.
35. A. E. Douglas and G. Herzberg, Can. J. Phys., vol. 29 (1951), 294.
36. G. Herzberg, Naturwiss., vol. 20 (1932), 577.
37. G. Herzberg, Can. J. Phys., vol. 30 (1952), 185.
38. J. Dufay, C. R., Paris, vol. 213 (1941), 284.
39. J. W. Ellis and H. O. Kneser, Z. Physik, vol. 86 (1933), 583.
40. O. R. Wulf, Proc. Nat. Acad. Sci., vol. 14 (1928), 609.
41. W. Finkelnburg and W. Steiner, Z. Physik, vol. 79 (1932), 69.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLVI : SERIES III : JUNE, 1952
SECTION THREE

Dominated Convergence of Kantorovitch Polynomials in the Space L^p

P. L. BUTZER

Presented by R. L. JEFFERY, F.R.S.C.

1. Introduction

If $f(x)$ belongs to the class L^1, of Lebesque integrable functions over the closed interval [0, 1], then the Kantorovitch polynomial (see [3] or [4]) corresponding to the function $f(x)$ defined in [0, 1] is

$$P_n^f(x) = \sum_{\nu=0}^{n} (n+1)p_{\nu,n}(x) \int_{\nu/(n+1)}^{\nu+1/(n+1)} f(t)\,dt, \tag{1}$$

where

$$p_{\nu,n}(x) = \binom{n}{\nu} x^{\nu}(1-x)^{n-\nu}.$$

In the form of a singular integral, this polynomial can be written

$$P_n^f(x) = \int_0^1 K_n(x,t) f(t)\,dt \tag{2}$$

with non-negative kernel

$$K_n(x,t) = (n+1)p_{\nu,n}(x) \tag{3}$$

for $\frac{\nu}{n+1} < t \leqslant \frac{\nu+1}{n+1}$, $\nu = 0, 1, 2, \ldots, n; K_n(x, 0) = 0; 0 \leqslant x \leqslant 1$.

Let

$$F(x) = \int_0^x f(t)\,dt.$$

At every point x, where $f(x) = F'(x)$, that is, almost everywhere in [0, 1], Lorentz [4] has proved that

$$\lim_{n\to\infty} P_n^f(x) = f(x). \tag{4}$$

In §2 we obtain an independent proof of this result as an immediate application of a general theorem. In §3 we shall show that the polynomials $P_n^f(x)$ approach the corresponding function $f(x)$ dominatedly (for this term, in the space L^1, see [2, p. 61]) in the space L^p, $p > 1$. As a corollary of this fundamental theorem we deduce that the $P_n^f(x)$ "converge strongly" to $f(x)$; the proof of the latter result previously was based on a rather particular theorem (cf. [4]).

I with to thank Professor G. G. Lorentz and Professor W. J. Webber for their kind interest and helpful suggestions. This paper was written during the tenure of a grant from the National Research Council, Ottawa.

2. Convergence of the $P_n^f(x)$

Lemma 1. *If the non-negative kernel* $\chi_n(x, t)$, $0 \leqslant x \leqslant 1$, *as a function of t increases for* $t < x$ *and decreases for* $t > x$, *then the conditions*

$$\lim_{n\to\infty} \int_0^1 \chi_n(x, t)\, dt = 1, \tag{5}$$

$$\chi_n(x, t) \to 0 \text{ as } n \to \infty,\ x \neq t, \tag{6}$$

imply that

$$\int_0^1 \chi_n(x, t) f(t)\, dt \to f(x) \tag{7}$$

at all points where $f(x)$ *is the derivative of its indefinite integral.*

For the proof, we refer the reader to the work of Romanovsky [6].

We shall now show that our kernel (3) satisfies the hypothesis of the lemma. Obviously

$$\int_0^1 K_n(x, t)\, dt = \sum_{\nu=0}^{n} p_{\nu,n}(x) = 1. \tag{8}$$

Now if $\delta = n^{-\alpha}$, $0 < \alpha < \frac{1}{2}$, then it is known that (cf. [4]) for every $k > 0$ there exists a constant C depending only on α and k such that

$$\sum_{|\nu/n-x} p_{\nu,n}(x) \leqslant Cn^{-k}.$$

From this inequality we obtain $(n + 1)p_{\nu,n}(x) \to 0$ for $|\nu/n - x| \geqslant \delta > 0$, so $K_n(x, t) \to 0$ for $t \neq x$. Furthermore as

$$\frac{p_{\nu,n}(x)}{p_{\nu-1,n}(x)} = \frac{x}{1-x} \cdot \frac{n - \nu + 1}{\nu},$$

$p_{\nu,n}(x)$ as a function of ν increases with ν for fixed n and x as long as $\nu/n + 1 \leqslant x$ and then decreases for larger ν. Hence $K_n(x, t)$ increases for $t < x$ and decreases for $t > x$. From (7) we obtain $P_n^f(x) \to f(x)$ p.p. (that is, almost everywhere).

3. DOMINATED CONVERGENCE OF THE $P_n'(x)$ IN THE SPACE L^p

First of all, we need the inequality

$$(9)\quad \sum_{\nu=0}^{n-1}(n+1)\left|\frac{\nu+1}{n+1}-x\right|\,\left|p_{\nu,n}(x)-p_{\nu+1,n}(x)\right|\leqslant 1,\qquad 0\leqslant x\leqslant 1.$$

Since

$$p_{\nu,n}(x)-p_{\nu+1,n}(x)=p_{\nu,n}(x)\frac{(\nu+1)-(n+1)x}{(\nu+1)(1-x)}$$

$$=\frac{1}{x}\frac{\nu+1}{n+1}\cdot\frac{(\nu+1)-(n+1)x}{(\nu+1)(1-x)}\cdot p_{\nu+1,n-1}(x),$$

the left-hand side of (9) equals

$$\frac{1}{x(1-x)}\sum_{\nu=0}^{n-1}\left|\frac{(\nu+1)-(n+1)x}{n+1}\right|\,\left|\frac{(\nu+1)-(n+1)x}{\nu+1}\right|\cdot(\nu+1)p_{\nu+1,n+1}(x)$$

$$=\frac{1}{x(1-x)}\sum_{\nu=0}^{n-1}[(\nu+1)-(n+1)x]^2\frac{p_{\nu+1,n+1}(x)}{n+1}\leqslant 1$$

because it is known that (see [4] or [7], p. 152])

$$\sum_{\nu=0}^{n}(\nu-nx)^2p_{\nu,n}(x)=nx(1-x).$$

This establishes the inequality (9).

We shall also make use of the following general lemma.

LEMMA 2. *Let* $\chi_n(x,t)$, *fixed* $0\leqslant x\leqslant 1$ *be a non-negative function of bounded variation in the variable* t, $0\leqslant t\leqslant 1$, *depending on the parameter* n, *with* $\chi_n(x,0)=0$ *and satisfying the conditions*

$$(10)\qquad \int_0^1\chi_n(x,t)\,dt\leqslant C,$$

$$(11)\qquad \int_0^1|t-x|\cdot|d_t\chi_n(x,t)|\leqslant C_1,$$

where C *and* C_1 *are independent of* n. *Define*

$$H_n(x)=\int_0^1\chi_n(x,t)f(t)\,dt$$

where $f(t)\in L^1[0,1)$. *Then*

$$\sup|H_n(x)|\leqslant A\cdot\theta(x;f)\ \text{p.p.},$$

where the constant $A\,(\equiv C+2C_1)$ *is independent of* f *and* n *and*

$$\theta(x; f) = \sup_{\substack{0 \leqslant t \leqslant 1 \\ t \neq x}} \frac{1}{t - x} \int_x^t |f(u)|\, du.$$

The proof of this lemma follows by a method similar to the one in [8, pp. 246–7].

The hypothesis of this lemma is readily seen to be satisfied by our kernel $K_n(x, t)$. In particular, the relation (8) shows that the inequality (10) is satisfied with $C = 1$ and on account of (9), (11) is satisfied with $C_1 = 1$. Hence we obtain that the polynomials $P_n{}^f(x)$ are majorized by the function $3\theta(x; f)$ for every n ($n = 0, 1, 2, \ldots$), that is,

$$\sup_n |P_n{}^f(x)| \leqslant 3\,\theta(x; f). \tag{12}$$

Now Hardy and Littlewood (see [8, p. 244]) have shown that if $f \in L^p[0, 1]$, $p > 1$ then $\theta(f) \in L^p[0, 1]$ with

$$\int_0^1 \theta^p(x; f)\, dx \leqslant 2\left(\frac{p}{p-1}\right)^p \int_0^1 |f(x)|^p\, dx. \tag{13}$$

In view of the relations (4), (12), and (13), we deduce the fundamental theorem.

THEOREM. *If* $f \in L^p[0, 1]$, $p > 1$, *the polynomials* $P_n{}^f(x)$ *are dominatedly convergent almost everywhere in the space* L^p, *in other words*

(i) $P_n{}^f(x) \rightarrow f(x)$ p.p. *in* $[0, 1]$,

(ii) *there is an element* $\theta(x; f)$ *in the space* L^p *such that* $|P_n{}^f(x)| \leqslant 3\theta(x; f)$ p.p., *for all* n.

COROLLARY. *If* $f \in L^p[0, 1]$, $p > 1$, *the polynomials* $P_n{}^f(x)$ *converge strongly to* $f(x)$, *namely*

$$||P_n{}^f - f||_{L^p} \equiv \left[\int_0^1 |P_n{}^f(x) - f(x)|^p\, dx\right]^{1/p} \rightarrow 0. \tag{14}$$

For, $|P_n{}^f(x) - f(x)|^p$ approaches zero almost everywhere and is dominated by an integrable function, and hence by Lebesgue's general convergence theorem the relation (14) follows.

The same methods of proof have been used to establish similar theorems in the Banach spaces $\Lambda(\alpha, p)$ and $M(\alpha, p)$, $0 < \alpha < 1$, $p > 1$, introduced by Lorentz [5]. The details are found in [1].

Recently, E. Levi [Atti. Accad. Naz. Lincei. Rend. Cl. Sci. Fis. Mat. Nat. (8), vol. 9 (1950), 242-246; (8), vol. 10 (1951), 360-364] has also established the above corollary proving that the $P_n{}^f(x)$ converge strongly to $f(x)$ in the space L^p, $p \geqslant 1$. His proof involves approximations by step functions.

REFERENCES

1. P. L. Butzer, *On Bernstein polynomials*, Thesis, University of Toronto Library, 1951.
2. G. H. Hardy and W. W. Rogosinski, *Fourier series* (Cambridge Tracts No. 38, 1944).
3. L. V. Kantorovitch, *Sur certaines développements suivant les polynomes de la forme de S. Bernstein*, I, II, C. R. Acad. Sci. U.R.S.S., A (1930), 563–8, 595–600.
4. G. G. Lorentz, *Bernstein polynomials*, a book to be published by the University of Toronto Press.
5. ——— *Some new functional spaces*, Ann. Math., vol. 51 (1950), 37–55.
6. M. P. Romanovsky, *Quelques considérations sur la théorie des integrales singulières*, Math. Z., vol. 34 (1932), 34–49.
7. D. V. Widder, *The Laplace transform* (Princeton, 1946)
8. A. Zygmund, *Trigonometrical series* (Warsaw, 1935).

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLVI : SERIES III : JUNE, 1952
SECTION THREE

On the Relation between the P^2-integral and the Cesàro-Perron Scale of Integrals

H W ELLIS

Presented by R L JEFFERY, F R S.C.

1. R. D. James [3] has defined a second Perron integral (P^2-integral) and shown that it contains and is more general than the Cesàro-Perron integral of J. C. Burkill. The Cesàro-Perron integral is the integral of order one in the Cesàro-Perron scale [1] of integrals of increasing generality. The purpose of this note is to give examples showing that there are functions that are P^2-integrable without being C_rP-integrable for any r and that, for r an integer greater than one, there are functions that are C_rP-integrable without being P^2-integrable.

2. *The functions*

$$f(p, x) = x^{-p}\sin x^{-q},\ g(p, x) = x^{-p}\cos x^{-q}, \quad p > 0,\ q > 0,\ x \neq 0;$$

$$f(p, 0) = g(p, 0) = 0$$

are C_rP-integrable on any finite interval containing the origin if and only if $r > (p - q - 1)/q$.

Fix $2k > (p - q - 1)/q$. The $2k$th derivative $f^{(2k)}[p - 2k(q + 1), x]$ exists for $x \neq 0$ and consists of a linear combination of the functions $f(p, x)$, $g(p - q, x)$, . . . , $f[p - 2kq, x]$. Defining

$$F_r(x) = f^{(2k-r)}[p - 2k(q + 1), x], \qquad 0 \leqslant r \leqslant 2k,\ x \neq 0;$$

$$F_r(0) = 0, \qquad f(x) = F_0(x)$$

gives

$$F_{r+1}(x) = 0[x^{(r+1)(q+1)-p}] \qquad \text{as } x \to 0. \tag{1}$$

The function $f(x)$ will be C_rP-integrable on (a, b) (with $F_1(x) - F_1(a)$ serving as both a major and minor function) if $F_1(x)$ is C_r-continuous at $x = 0$, that is if

$$\lim_{x\to 0} \frac{r}{x^r}\int_0^x (x - t)^{r-1}F_1(t)\,dt = F_1(0) = 0,$$

the integration being in the $C_{r-1}P$-sense. By successive integrations by parts, valid for the Cesàro-Perron integrals involved, this is equivalent to

$$(2) \qquad F_{r+1}(x) = 0(x^r) \qquad \text{as } x \to 0.$$

It follows that $f(x)$ is C_rP-integrable on (a, b) if $r > (p - q - 1)/q$.

Similarly $f_1(x) = g^{(2k)}[p - 2k(q + 1) - q, x]$, $x \neq 0$, $f_1(0) = 0$ is a linear combination of the functions $g(p - q, x)$, $f(p - 2q, x)$, . . . , $g[p - (2k + 1)q, x]$ and is C_rP-integrable on (a, b). Continuing in this way a C_rP-integrable linear combination $f_n(x)$ of the functions $f(p - nq, x)$, . . . , $f[p - (2k + n)q, x]$ is obtained for n even and a similar combination with f and g interchanged for n odd. For the smallest integer $n > (p + 1)\ q$, all of the terms of $f_n(x)$ except $f(p - nq, x)$ (n even) or $g(p - nq, x)$ (n odd) will be Lebesgue integrable and therefore $f(p - nq, x)$ or $g(p - nq, x)$ will be C_rP-integrable. Reversing the order in which the above linear combinations were formed, the C_rP-integrability of the function of lowest order follows at each step and finally the C_rP-integrability of $f(p, x)$ is obtained.

A consideration of (1) and (2) shows that $f(x)$ is not C_rP-integrable if $r \leqslant (p - q - 1)/q$. We may suppose that $r + 1 > (p - q - 1)/q$, for if it is shown that $f(p, x)$ is not C_rP-integrable for such a value of r it follows from the consistency of the Cesàro-Perron scale of integrals that $f(p, x)$ cannot be $C_{r'}P$-integrable for $r' < r$. Under this hypothesis $f_1(x)$ is C_rP-integrable as are all of the terms of $f(x)$ except $f(p,x)$. It follows that $f(p, x)$ cannot be C_rP-integrable on (a, b).

3. *For r fixed there are functions that are P^2-integrable without being C_rP-integrable.* Consider

$$F(x) = x^\alpha \sin x^{-\beta},\ G(x) = x^\alpha \cos x^{-\beta},\ x \neq 0,\ F(0) = G(0) = 0,$$

$$f(x) = F''(x),\ g(x) = G''(x),\ x \neq 0,\ f(0) = g(0) = 0.$$

By James's Theorem 4.2 [3] the functions $f(x)$, $g(x)$ will be P^2-integrable over (a, c, b) where (a, b) contains the origin if $F(x)$, $G(x)$ are continuous and smooth at $x = 0$. Continuity will be assured by (i) $\alpha > 0$; smoothness by (ii) x^β even, x^α odd for $F(x)$ or $G(x)$, or x^β odd, x^α even for $F(x)$ or both odd for $G(x)$. From §2, $f(x)$ will not be C_rP-integrable if (iii) $\alpha \leqslant \beta(1 - r) + 1$. Values of α and β satisfying (i)–(iii) include $\beta = 2(2r + 1)^{-1}$, $\alpha = 3(2r + 1)^{-1}$ in the definition of $F(x)$ or $G(x)$, or $\beta = (2r + 1)^{-1}$, $\alpha = (r + 2)/(2r + 1)$ in the definition of $F(x)$ if r is even and of $G(x)$ if r is odd.

4. *There are functions that are P^2-integrable without being C_rP-integrable for any r.*

Let I_n denote the interval $[2^{-n}, 2^{-n+1}]$ and define

$$F_n(x) = (x - x_n)^{3/(2n+1)} \cos \frac{C_n}{(x - x_n)^{2/(2n+1)}}; \qquad x, x_n \in I_n,$$

$$F(x) = -F(-x) = \sum_{n=1}^{\infty} a_n F_n(x),$$

$$f(x) = F''(x),\ x \neq 0,\ x \neq x_n;\ n = 1, 2, \ldots;\ f(x) = 0 \text{ elsewhere.}$$

It follows from §3 that $f(x)$ is not C_rP-integrable on $(-1, 1)$ for any r. We shall show that the constants x_n, C_n and $a_n (n = 1, 2, \ldots)$ can be so chosen that $f(x)$ is P^2-integrable over $(-1, c, 1)$ for every $-1 < c < 1$.

$F_n(x)$ has maxima for

$$\left[\frac{C_n}{2k\pi}\right]^{n+\frac{1}{2}} < x - x_n < \Big| \frac{2C_n}{(4k-1)n}$$

and the values of $k = 1, 2, \ldots$ for which $x \in I_n$. The maximum of $F_n(x)$ corresponding to a fixed k (notation $\max_k F_n(x)$) satisfies

$$\left[\frac{C_n}{2k\pi}\right]^{3/2} < \max_k F_n(x) < \left[\frac{2C_n}{(4k-1)n}\right]^{3/2}.$$

If x_n and C_n are chosen so that $\max_1 F_n(x)$ is at $x = 2^{-n+1}$, $\max_{-p} F_n(x)$ at $x = 2^{-n}$ it may be verified that

$$\frac{\max_{-p} F_n(x)}{\max_1 F_n(x)} < \left[\frac{16}{3(4p-1)}\right]^{3/2}.$$

If $p \geqslant 2$ is fixed and the values a_n chosen so that

$$a_n F_n(2^{-n}) = a_{n+1} F_{n+1}(2^{-n}),$$

it then follows that, since $\max_1 a_{n+1} F_{n+1}(x) = \max_{-p} a_n F_n(x)$,

$$\frac{\max_1 a_{n+1} F_{n+1}(x)}{\max_1 a_n F_n(x)} < \lambda < 1, \text{ all } n.$$

Since $\max_1 a_n F_n(x)$ is the maximum of $F(x)$ on I_n, $F(x)$ is continuous at $x = 0$. Since $F(x)$ is odd and $F(0) = 0$, $F(x)$ is smooth at $x = 0$. By definition $F(x)$ is continuous at the points $\pm 2^{-n}$, $n = 0, 1, 2, \ldots$ and is smooth at these points since both right and left derivatives are zero there. As in §3, $F(x)$ is smooth and continuous at each point x_n and the P^2-integrability of $f(x)$ follows from James's Theorem 4.2.

5. If $\alpha = 0$ in the definition of $g(x)$ in §3, $g(x)$ is not P^2-integrable over $(-1, c, 1)$ since $F(x)$ cannot be defined so as to be continuous at $x = 0$. By §2 $g(x)$ is C_rP-integrable if $r > (\beta + 1)/\beta$ and is therefore C_2P-integrable for $\beta > 1$.

All of the preceding relations hold with the more genreal GM_r-integrals [2] replacing the corresponding C_rP-integrals.

REFERENCES

1. J. C. Burkill, *The Cesàro-Perron scale of integration*, Proc. London Math. Soc. (2), vol. 39 (1935), 541-52.
2. H. W. Ellis, *Mean-continuous integrals*, Can. J. Math., vol. 1 (1949), 113-24.
3. R. D. James, *A generalised integral* II, Can. J. Math., vol. 2 (1950), 297-306.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLVI : SERIES III : JUNE, 1952
SECTION THREE

Unbounded Operators and a Theorem of A. Robinson*

G. G. LORENTZ and M. S. MACPHAIL

Presented by R. L. JEFFERY, F.R.S C.

1. INTRODUCTION

LET B be a Banach space of elements x with convergence defined as usual by the norm. Let t_{mn} $(m, n = 1, 2, \ldots)$ be linear (not necessarily bounded) operators mapping B into itself. In a recent paper, A. Robinson [4] (see also Melvin-Melvin [3]) has proved the following interesting generalization of the theorem of Toeplitz on regular summability methods. The word "bounded," applied to an operator T (but not to its value Tx) will always mean the existence of the norm $\|T\|$.

THEOREM 1. *In order that the transformation*

$$y_m = \sum_{n=1}^{\infty} t_{mn} x_n \tag{1}$$

be defined for each convergent sequence $x_n \in B$ *and that* $\lim y_m = \lim x_n$ *it is necessary and sufficient that*

$$\lim_{m\to\infty} t_{mn} = 0 \qquad (n = 1, 2, \ldots), \tag{2}$$

$$\sum_{n=1}^{\infty} t_{mn} = t_m \text{ exists for } m = 1, 2, \ldots, \tag{3}$$

$$\lim_{m\to\infty} t_m = I, \tag{4}$$

$$\|t_{mN}, t_{m,N+1}, \ldots\| \leqslant M < +\infty \qquad (m = 1, 2, \ldots) \tag{5}$$

for some constants M *and* N.

Here 0, I are the zero and the identical operator, respectively, and (2) means for instance that $t_{mn}\,x \to 0$, $m \to \infty$ for each $x \in B$. The "group norm" $\|t_\nu, t_{\nu+1}, \ldots\|$ is defined by

$$\|t_\nu, t_{\nu+1}, \ldots\| = \sup \|t_\nu x_\nu + \ldots + t_\mu x_\mu\| \tag{6}$$

taken over all integers $\mu \geqslant \nu$ and all x_p with $\|x_p\| \leqslant 1$ $(p = \nu, \ldots, \mu)$.

---

*This paper was written while the authors were members of the Summer Research Institute, Canadian Mathematical Congress, Kingston, 1952.

Condition (5) implies that the operators t_{mn} must be bounded for $n \geqslant N$. If they are assumed bounded for all n, $m = 1, 2, \dots$, this theorem follows easily from the Banach-Steinhaus theorem for operators on $c(B)$ (see §2); Robinson and Melvin-Melvin give much longer proofs. The purpose of this note is to give a general theorem on linear (possibly unbounded) operators which seems to be new (Theorem 3 in §3 and its variant, Theorem 4) and to show that the general case of Theorem 1 follows from the special case and this result. The existence of such a theorem was suggested to us by Robinson's general theorem.

2. BOUNDED OPERATORS

For a given Banach space B, we denote by $c(B)$ the set of all sequences $X = \{x_n\}$, $x_n \in B$, converging in B. $c(B)$ is a Banach space with the norm

$$||X|| = \sup_n ||x_n||.$$

If the t_ν $(\nu = 1, \dots, n)$ are linear operators from B to B, the sum

$$\sum_{\nu=1}^{n} t_\nu x_\nu = TX$$

is a linear operator from $c(B)$ to B; its norm is easily seen to be

$$||T|| = ||t_1, \dots, t_n|| = \sup_{||x_\nu|| \leqslant 1} ||\sum_1^n t_\nu x_\nu||.$$

For a sequence $t_1, \dots, t_n, \dots$, $||t_1, \dots, t_n|| \to ||t_1, t_2, \dots||$. A fundamental set in $c(B)$ consists of all sequences of the form $(0, \dots, 0, x, 0, \dots)$ and $(x, x, \dots)$. Hence, by the Banach-Steinhaus theorem [1, p. 80, theorem 5] a series $\sum_1^\infty t_\nu x_\nu$ is a bounded linear operator TX from $c(B)$ to B if and only if $\sum t_n$ converges and $||t_1, t_2, \dots|| < +\infty$ (this is Theorem III in [4]). Moreover,

$$||T|| = ||t_1, t_2, \dots||.$$

Assume now that the t_{mn} in (1) are bounded operators. Applying the Banach-Steinhaus theorem to the sequence of operators

$$T_m X = \sum_n t_{mn} x_n, \quad X = \{x_n\},$$

we obtain immediately Theorem 1 for our special case.

[Remark during the proof (December 19, 1952): essentially the same proof was given by K. Zeller in his paper [*Verallgemeinerte Matrixtransformationen*, Math. Z., vol. 56 (1952), 18 20] which appeared after this note was accepted for publication. However, Zeller leaves open the case of Robinson's general theorem.]

We note also the following theorem, whose proof follows along the same lines as [2, theorem 1].

THEOREM 2. *The transformation* (1) *has the property* $\sum \|y_m\| < +\infty$ *whenever* $\sum \|x_n\| < +\infty$ *if and only if, for some constant* M,

$$\sum_{m=1}^{\infty} \|t_{mn}\, x\| \leqslant M \quad \text{for } \|x\| \leqslant 1, \qquad n = 1, 2, \ldots . \tag{7}$$

3. UNBOUNDED OPERATORS

We now give the theorems referred to in the introduction.

THEOREM 3. *Let* B *be a Banach space and* L_m $(m = 1, 2, \ldots)$ *closed linear subspaces of* B. *Let* T_m *be a linear operator mapping* B *into itself, such that* T_m *is bounded on* L_m. *Let* $T_1 x, T_2 x, \ldots$ *be a bounded sequence for each* $x \in B$. *Then there is an* N *such that all* T_m *are bounded on* $\bigcap_{m=1}^{N} L_m$.

Proof. We may assume without loss of generality that $L_1 \supset L_2 \supset \ldots$, for otherwise we should consider

$$L_m^{*} = \bigcap_{\mu=1}^{m} L_\mu .$$

Suppose the conclusion does not hold. Then for each L_m there is a T_n which is not bounded on L_m, where necessarily $n > m$. Passing to a subsequence of the T_m, we may assume T_{m+1} is not bounded on L_m, for each m.

Let H_m be the bound of T_m on L_m. Choose a sequence $\alpha_1, \alpha_2, \ldots$ with

$$H_{m+1} \sum_{\mu > m} \alpha_\mu \leqslant 1. \tag{8}$$

Choose a sequence $x_1, x_2, \ldots$ of elements of B inductively, as follows. If $x_1, \ldots, x_{m-1}$ are already known, choose x_m such that $x_m \in L_m$, $\|x_m\| \leqslant 1$, and

$$\alpha_m \|T_{m+1}\, x_m\| > m + \sum_{\mu=1}^{m-1} \alpha_\mu \|T_{m+1}\, x_\mu\|. \tag{9}$$

We put

$$x = \sum_{\mu=1}^{\infty} \alpha_\mu\, x_\mu \in B;$$

then

$$T_{m+1}\, x = \alpha_n\, T_{m+1}\, x_m + \sum_{\mu=1}^{m-1} \alpha_\mu\, T_{m+1}\, x_\mu + T_{m+1} \sum_{\mu=m+1}^{\infty} \alpha_\mu\, x_\mu. \tag{10}$$

Here $x_\mu \in L_{m+1}$ for $\mu \geqslant m+1$, and so

$$\sum_{m+1}^{\infty} \alpha_\mu\, x_\mu \in L_{m+1}.$$

Hence

$$(11) \qquad ||T_{m+1} \sum_{\mu > m} \alpha_\mu x_\mu|| \leqslant H_{m+1} \sum_{\mu > m} \alpha_\mu \leqslant 1.$$

From the relations (10), (9), (11), (8) it follows that $\{T_m x\}$ is unbounded. This contradiction proves the theorem.

THEOREM 4. *Let $L_1 \subset L_2 \subset \ldots$ be closed linear subspaces of a Banach space B. Let $T_m = S_m + T_m'$ $(m = 1, 2, \ldots)$, where T_m, S_m, T_m' are linear operators mapping B into itself, each T_m' is bounded, and S_m vanishes outside of L_m. Let $\{T_m x\}$ be bounded for each $x \in B$. Then there is an N such that all T_m are bounded on the part of the unit sphere not contained in L_N.*

The proof is a variation of that of Theorem 3 Let U be the unit sphere of B. We may assume that T_{m+1} is not bounded on $U - L_m$. We take $H_m = ||T_m'||$, and define the α_m as before. The x_m are defined inductively such that

$$x_m \in U - L_m, \quad \alpha_m ||T_{m+1} x_m|| > m + \sum_{\mu=1}^{m-1} \alpha_\mu ||T_{m+1} x_\mu||,$$

$$\alpha_m ||x_m|| < \max \left\{ \frac{1}{3} d_{m-1}, \frac{1}{9} d_{m-2}, \ldots, \frac{1}{3^{m-1}} d_1 \right\},$$

where $d_\mu = \rho(\alpha_\mu x_\mu, L_\mu) > 0$ is the distance from $\alpha_\mu x_\mu$ to L_μ.

Then for

$$z_m = \sum_{\mu=m+1}^{\infty} \alpha_\mu x_\mu ,$$

$$\rho(z_m, L_{m+1}) \geqslant d_{m+1} - \sum_{\mu=m+2}^{\infty} \alpha_\mu ||x_\mu|| \geqslant d_{m+1} - \frac{1}{3} d_{m+1} - \frac{1}{9} d_{m+1} - \ldots$$

$$= \frac{1}{2} d_{m+1} > 0.$$

This shows that $z_m \notin L_m$. Therefore

$$||T_{m+1} z_m|| = ||T_{m+1}' z_m|| \leqslant H_{m+1} \sum_{m+1}^{\infty} \alpha_\mu \leqslant 1,$$

and the proof is completed by applying (10).

Theorems 3 and 4 remain true even if T_m maps B into a Banach space B_m depending on m. A similar remark applies to Theorems 1 and 2

4 THE PROOF OF THEOREM 1

We begin with the following remark. Let $t_1, t_2, \ldots$ be a sequence of linear operators, mapping B into itself. If $\{t_n x_n\}$ is a bounded sequence for each $\{x_n\} = X \in c(B)$, then there is a number ν such that each operator t_n with $n > \nu$ is bounded For otherwise there would be indices $n_1 < n_2 < \ldots$ and elements x_{n_l} with

$$\|x_{n_l}\| \leqslant 1 \quad \text{and} \quad \|t_{n_l} x_{n_l}\| \to +\infty.$$

Then if

$$y_l = x_{n_l} \|t_{n_l} x_{n_l}\|^{-\frac{1}{2}},$$

we should have

$$y_l \to 0 \quad \text{and} \quad \|t_{n_l} y_l\| \to +\infty,$$

a contradiction. This proves our statement.

If now an operator of the form $T_m X = \sum_n t_{mn} x_n$ is defined for each $X \in c(B)$, it follows that there is a closed linear subspace L_m of $c(B)$, viz , the set of sequences with $x_1 = x_2 = \ldots = x_\nu = 0$, on which T_m is bounded. Hence, by Theorem 3, there is an N such that each T_m ($m = 1, 2, \ldots$) is bounded on the set of sequences with $x_1 = x_2 = \ldots = x_N = 0$. Clearly this implies that all t_{mn} are bounded for $n > N$. By combining this with the special case of Theorem 1 proved in §2, we obtain the complete proof of Theorem 1

REFERENCES

1. S Banach, *Théorie des opérations linéaires* (Warsaw, 1932)
2. K. Knopp and G. G Lorentz, *Beiträge zur absoluten Limitierung*, Archiv der Mathematik, vol 2 (1949), 10–16
3. H. Melvin-Melvin, *On generalized K-transformations in Banach spaces*, Proc. London Math Soc (2), vol 53 (1951), 83–108
4. A Robinson, *On functional transformations and summability*, Proc London Math. Soc (2), vol 52 (1950), 132–160

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLVI SERIES III . JUNE, 1952
SECTION THREE

The Extended Euler-Knopp Transformation

M. S. MACPHAIL

Presented by R. L. JEFFERY, F.R.S.C.

Summation methods based on the following transformations

$$G_1(p): \qquad b_n = \sum_{k=n}^{\infty} \binom{k}{n} p^{k-n}(1-p)^n a_k,$$

$$G_2(p): \qquad b_n = \sum_{k=0}^{\infty} \binom{k+n-1}{n} p^k (1-p)^n a_k,$$

have recently been studied by W. Meyer-König [3] and P. Vermes [6]. Here p is a real or complex constant, and the series $\sum a_k$ is said to be $G_j(p)$-summable to s $(j = 1, 2)$ if the series $\sum b_n$ exists and converges to s. It was noted in [3] that these methods may be generated by the functions

$$g_1(t) = p + (1-p)t,$$

$$g_2(t) = p/\{1-(1-p)t\},$$

respectively, by arranging $\sum_k a_k\{g_j(t)\}^k$ formally in powers of t and putting $t = 1$; this yields the series $\sum b_n$ in each case. If we take

$$g_3(t) = pt/\{1-(1-p)t\}$$

we obtain the method

$$G_3(p): \quad b_0 = a_0, \; b_n = \sum_{k=1}^{n} \binom{n-1}{k-1} p^k (1-p)^{n-k} a_k \qquad (n = 1, 2, \ldots)$$

which can be shown by summing and differencing to be identical with the Euler-Knopp sequence-to-sequence method

$$E(p): \qquad \sigma_n = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} s_k.$$

Thus $G_1(p)$, $G_2(p)$, $G_3(p)$ are special cases of the method G in which we take any suitable function

$$g(t) = \sum_{n=0} c_n t^n,$$

with $g(1) = 1$, let $\sum a_k\{g(t)\}^k = \sum b_n t^n$ formally, and call $\sum a_k$ G-summable to s if $\sum b_n = s$. If we put

$$\{g(t)\}^k = \sum_{n=0}^{\infty} c_n^{(k)} t^n$$

(so that $c_n = c_n^{(1)}$), we may define the series-to-series method G by the transform

$$G: \qquad b_n = \sum_{k=0}^{\infty} c_n^{(k)} a_k$$

If $c_0 = 0$ we have a row-finite matrix; otherwise the rows are infinite. The case $c_0 = 0$, $c_n \geqslant 0$ $(n = 1, 2, \dots)$ was studied by O. Perron [4]; and the same with $c_n \geqslant 0$ relaxed, by the author [2]. K. Knopp [1] allowed $c_0 \neq 0$, and determined the w-region in which the series $\sum a_n w^n$ is G-summable, under certain conditions on $g(t)$, knowledge of the singularities of the function defined by $\sum a_n w^n$ being assumed (the analogue of the Borel polygon theorem).

The cases $c_0 = 0$ and $c_0 \neq 0$ require somewhat different treatment. The aim of the present note is to continue the study of the case $c_0 \neq 0$, giving the new form of some of the theorems in [2], and pointing out the application to the methods $G_1(p)$ and $G_2(p)$. Let us write out more fully the rearrangement involved in the method G:

$$\begin{array}{lrcl} & a_0 & = & a_0 \\ & a_1\, g(t) & = & a_1\, c_0^{(1)} + a_1\, c_1^{(1)} t + a_2\, c_2^{(1)} t^2 + \dots \\ (A) & a_2\{g(t)\}^2 & = & a_2\, c_0^{(2)} + a_2\, c_1^{(2)} t + a_2\, c_2^{(2)} t^2 + \dots \\ & & & \\ & \sum a_i\{g(t)\}^k & = & b_0 \qquad + b_1\, t \qquad + b_2\, t^2 \qquad + \dots \end{array}$$

We see at once that the assumption that b_n is defined for each n implies that each column forms a convergent series, and in particular, since $c_0^{(k)} = c_0^k$, the series $\sum a_k z^k$ has radius of convergence at least $|c_0|$. Furthermore, the assumption that $\sum b_n$ converges implies that $\sum b_n t^n$ has radius of convergence at least 1. We shall also assume that $g(1) = 1$, that the series $\sum c_n t^n$ has radius of convergence greater than 1, and that $z = g(t)$ maps $|t| = 1$ into a simple closed Jordan curve J in the z plane, and maps $|t| < 1$ simply on D, the interior of J.

THEOREM 1. *Let $\sum a_n z^n$ have radius of convergence $r > |c_0|$. Then if $\sum a_n$ is G-summable, the function $\psi(z) = \sum a_n z^n$ can be continued analytically through D, and conversely if $\psi(z)$ can be continued analytically through D′, the closure of D, then $\sum a_n$ is G-summable.*

Remark. This can be sharpened somewhat (see [3, theorem 4] for $G_1(p)$) but the present form is sufficient for our purposes.

Proof. It follows from our hypothesis by the Weierstrass double-series theorem that each column of (A) converges, and if $\phi(t) = \psi\{g(t)\}$ the expansion $\phi(t) = b_0 + b_1 t + \ldots$ is valid at least for sufficiently small values of t. If $\sum a_n$ is G-summable, $\phi(t)$ is defined for $|t| < 1$; then, if $t = h(z)$, the function $\phi\{h(z)\}$ gives the continuation of $\psi(z)$ through D. Conversely, if $\psi(z)$ can be continued analytically through D', then $\phi(t)$ is analytic on $|t| \leqslant 1$, and $\sum b_n t^n$ converges when $t = 1$. This completes the proof.

As an example, if $a_n = w^n$ the function $\psi(z) = \sum w^n z^n$ has radius of convergence $|1/w|$ and a pole at $z = 1/w$. Thus for $\sum w^n$ to be G-summable it is sufficient that $|1/w| > |c_0|$ and that $1/w$ does not lie in D', that is, $|h(1/w)| > 1$. Necessary conditions are $|1/w| > |c_0|$ and $|h(1/w)| \geqslant 1$; whether $|h(1/w)| > 1$ is necessary depends on G. For $G_1(p)$ and $G_2(p)$ see Vermes [6, pp. 550, 558]. For the summability of $\sum a_n w^n$ in the case where J contains the origin and is cut by each half-ray from the origin in one point only, see Knopp [1].

The method G is regular if all $c_n \geqslant 0$; the proof given by Perron [4, theorem 1] with $c_0 = 0$ is valid for $c_0 \neq 0$. If not all $c_n \geqslant 0$, the method G need not be regular; it may however be regular over certain classes of series as shown by Theorems 2, 2a, 3 that follow.

THEOREM 2. *If $R > |c_0|$, a necessary and sufficient condition for G to have the property that $\sum a_n$ is G-summable whenever $\sum a_n z^n$ has its radius of convergence greater than R, is that D' should lie in the circle $|z| \leqslant R$.*

This follows from Theorem 1. (We remark that if $R < 1$ the theorem is vacuous since J passes through $z = 1$.) More generally, if K is any closed domain containing the circle $|z| \leqslant c_0$ in its interior, we have

THEOREM 2a. *A necessary and sufficient condition for G to have the property that $\sum a_n$ is G-summable whenever $\sum a_n z^n$ represents a function analytic in K, is that D' should lie in K.*

This contains the main part of a theorem given by J. Teghem [5] for Euler-Knopp summability.

THEOREM 3. *Assume $|c_0| < 1$. Then G is consistent with convergence; that is, if $\sum a_n$ and $\sum b_n$ both converge, they converge to the same value.*

Proof. If $\sum a_n z^n$ converges for $z = 1$, its radius of convergence is greater than $|c_0|$, so the considerations of Theorem 1 apply. The result follows by Abel's theorem on power series. Similarly in Theorem 2, $\sum a_n$ is G-summable to its ordinary value.

THEOREM 4. *Let $\sum a_n$ be G-summable, and let $\sum a_n z_n$ have radius of convergence greater than $|c_0|$. Then $\sum a_n w^n$ is G-summable for all values of w such that $wD' \subset D$.*

Proof. We write $\Psi(z) = \sum a_n w^n z^n$. By Theorem 1, $\sum a_n w^n$ is G-summable if $\sum a_n w^n z^n$ has radius of convergence greater than $|c_0|$ and $\Psi(z)$ is analytic in D'. Now $wD' \subset D$ implies $|w| < 1$, so the radius of convergence of $\sum a_n w^n z^n$ is greater than that of $\sum a_n z^n$, hence greater than $|c_0|$. Also $\Psi(z) = \psi(wz)$, and since ψ is known to be analytic in D, the result follows.

THEOREM 5. *Let the series $\sum a_n$, $\sum a_n'$ be G-summable to s, s', and let $\sum a_n z^n$, $\sum a_n' z^n$, have radii of convergence greater than $|c_0|$. Let $\sum a_n''$ be the Cauchy product, $a_n'' = a_0 a_n' + \ldots + a_n a_0'$. Then $\sum a_n''$ is G-summable to ss', provided $\sum b_n$, the G-transform of $\sum a_n$, is absolutely convergent.*

For if $\sum b_n'$, $\sum b_n''$ are the G-transforms of $\sum a_n'$, $\sum a_n''$, then $\sum b_n''$ is none other than the Cauchy product of $\sum b_n$, $\sum b_n'$. The result follows by Merten's theorem.

In conclusion we state briefly the results for $G_1(p)$ and $G_2(p)$. For $G_1(p)$ we have $c_0 = p$, and the region D is $|z - p| < |1 - p|$. If $\Re p < \frac{1}{2}$, D contains the origin and the shortest distance from J to the origin is $|1 - p| - |p|$. If $|p| < \frac{1}{2}|1 - p|$ we have $|1 - p| - |p| > |p|$, so the assumption that $\sum a_n z^n$ is analytic in D' implies that the radius of convergance of $\sum a_n z^n$ is greater than $|c_0|$; therefore $\sum a_n$ is $G_1(p)$-summable by Theorem 1. The condition in Theorem 2, that D' lies in $\leqslant R$, becomes

$$|1 - p| + |p| \leqslant R.$$

The condition in Theorem 4, that $wD' \subset D$, becomes

$$|p/(1 - p)|\,|w - 1| + |w| < 1.$$

For $G_2(p)$ we require $|1 - p| < 1$. We have again $c_0 = p$, and the region D is $|z - p| < |z|\,|1 - p|$. This is a circle with centre $p/\{1 - |1 - p|^2\}$ and radius $|1 - p|\,|p|/\{1 - |1 - p|^2\}$. The condition in Theorem 2, that D' lies in $|z| \leqslant R$, becomes

$$|p/R| + |1 - p| \leqslant 1,$$

and the condition in Theorem 4, that $wD' \subset D$, becomes

$$|1/(1 - p)|\,|w - 1| + |w| < 1.$$

The corresponding results for $G_3(p)$ are given in [2].

REFERENCES

1. K. KNOPP, *Uber Polynomentwicklungen*. Acta Math., vol. 47 (1926), 313–35.
2. M. S. MACPHAIL, *On Perron's extension of the Euler-Knopp summation method*. Trans. Roy. Soc. Can., Ser. 3, Sec. III, vol. 42 (1948), 43–9.
3. W. MEYER-KONIG, *Untersuchungen uber einige verwandte Limitierungsverfahren*, Math. Z., vol. 52 (1949), 257–304.
4. O. PERRON, *Uber eine Verallgemeinerung der Eulerschen Reihentransformation*, Math. Z., vol. 18 (1923), 157–62.
5. J. TEGHEM, *Généralisation d'un théorème de M. Macphail sur la sommabilité d'Euler-Knopp*, Mathesis, vol. 58 (1949), 53–7.
6. P. VERMES, *Series-to-series transformations and analytic continuation by matrix methods*, Amer. J. Math., vol. 71 (1949), 541–62.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLVI : SERIES III : JUNE, 1952
SECTION THREE

Representations of Positive Real Numbers by Infinite Sequences of Integers

N S MENDELSOHN

Presented by G. de B ROBINSON, F.R S C.

1. INTRODUCTION

OUR standard decimal notation is based on the following theorems of arithmetic:

1. Any positive integer N can be written uniquely in the form

$$N = a_0 + 10a_1 + 10^2a_2 + \ldots + 10^r a_r$$

where a_i are integers such that $0 \leqslant a_i \leqslant 9$, $i = 1, 2, 3, \ldots, r$. In other words, the integer N determines the number r and each of the a_i.

2. Any real number M, $0 < M < 1$, can be written uniquely in the form

$$M = \frac{c_1}{10} + \frac{c_2}{10^2} + \ldots + \frac{c_r}{10^r} + \ldots$$

where c_i are integers such that $0 \leqslant c_i \leqslant 9$ for $i = 1, 2, 3, \ldots$, and furthermore, for infinitely many i, $c_i \neq 9$. If we remove the latter part of the condition, uniqueness is destroyed, for example, the numbers 0.5000 . . . and 0.49999 . . . are identical.

These theorems are readily generalized to the case where the number 10 is replaced by the number a, where a is any integer $\geqslant 2$. Less trivial generalizations have been given in [2] by Faber who has shown that:

1. Every integer can be expressed uniquely in the form

$$b_1 + b_2\, 2! + b_3\, 3! + \ldots + b_n\, n!$$

where b_i are integers such that $0 \leqslant b_i \leqslant i$.

2. Every number M, $0 < M < 1$, can be expressed uniquely in the form

$$M = \frac{c_1}{2!} + \frac{c_2}{3!} + \frac{c_3}{4!} + \ldots + \frac{c_r}{(r+1)!} + \ldots$$

where c_i are integers such that $0 \leqslant c_i \leqslant i$ and for infinitely many i, $c_i < i$. Faber obtains, from this representation, a fascinating proof of the countability of the rational numbers. In [1] Cantor represents positive numbers less than one by expressions of the form

$$\frac{\lambda}{b} + \frac{\mu}{bb'} + \frac{\nu}{bb'b''} + \dots$$

where $\lambda, \mu, \nu, \dots$ are integers such that $0 \leqslant \lambda \leqslant b - 1, 0 \leqslant \mu \leqslant b' - 1$, etc. In this paper, we will show that, in a sense, Cantor's representation is the most general possible. Our purpose is actually to find all possible representations of real numbers by sequences of integers, of the type described in this introduction.

2. The Main Theorem

Let $\alpha_0 = 1, \alpha_1, \alpha_2, \dots, \alpha_n, \dots$ be a sequence of real numbers strictly increasing to infinity; that is $\alpha_{i+1} > \alpha_i$ for $i = 0, 1, 2, \dots$ and for any real number N, there is an integer n such that $\alpha_n > N$. Let $\beta_1, \beta_2, \dots, \beta_n, \dots$ be a second sequence of real numbers strictly increasing to infinity. Let $\lambda_1, \lambda_2, \dots, \lambda_n$ be any sequence of positive integers and let $\mu_1, \mu_2, \dots, \mu_n$ be any second sequence of positive integers, including the possibility that $\mu_i = \lambda_i$. We ask, what are the necessary and sufficient conditions in order that every positive real number N may be expressed uniquely in the form

$$N = a_n\,\alpha_n + a_{n-1}\,\alpha_{n-1} + \dots + a_0\,\alpha_0 + \frac{b_1}{\beta_1} + \dots + \frac{b_i}{\beta_i} + \dots$$

where $a_0, a_1, \dots, a_n, b_1, b_2, \dots$ are integers such that

$$0 \leqslant a_i \leqslant \lambda_i, \quad 0 \leqslant b_i \leqslant \mu_i$$

and such that for infinitely many i, $b_i < \mu_i$.

In order to simplify our notation we will replace μ_n by λ_{-n} and $1/\beta_n$ by α_{-n} when these are more convenient. Then $\alpha_r \to 0$ as $r \to -\infty$ and $\alpha_r \to \infty$ as $r \to \infty$. We will also use a reverse summation notation, i.e., we will use

$$*\sum_{i=r}^{s} u_i$$

for the sum

$$u_r + u_{r-1} + \dots + u_s.$$

This notation will only be used in the case where $r \geqslant s$ and the order of summation will only be important when $s = -\infty$.

THEOREM 1. *A necessary and sufficient condition that every real positive number N be expressible uniquely in the form*

$$N = {}^*\sum_{i=r}^{-\infty} a_i\,\alpha_i$$

where $0 \leqslant a_i \leqslant \lambda_i$ and for infinitely many i, $a_i < \lambda_i$ is that

$$\alpha_{k+1} = {}^*\sum_{i=k}^{-\infty} \lambda_i\,\alpha_i \qquad \textit{for } -\infty < k < \infty.$$

Proof of sufficiency. Assume

$$\alpha_{k+1} = {}^*\sum_{i=k}^{-\infty} \lambda_i\,\alpha_i \qquad \text{for } -\infty < k < \infty.$$

Hence

$$\alpha_0 = {}^*\sum_{i=-1}^{-\infty} \lambda_i\,\alpha_i = 1.$$

For $n > 0$,

$$\alpha_n = {}^*\sum_{i=n-1}^{0} \lambda_i\,\alpha_i + {}^*\sum_{i=-1}^{-\infty} \lambda_i\,\alpha_i = {}^*\sum_{i=n-1}^{0} \lambda_i\,\alpha_i + 1.$$

We first show that α_i is an integer for all $i \geqslant 0$. We have $\alpha_0 = 1$, an integer, and assuming $\alpha_0, \alpha_1, \ldots, \alpha_{n-1}$ are integers, the equation

$$\alpha_n = {}^*\sum_{i=n-1}^{0} \lambda_i\,a_i + 1$$

shows that α_n is also an integer since all the λ_i are integral. Now, let N be any positive real number. Put $N = K + L$ where K is an integer and $0 \leqslant L < 1$. We will represent K in the form

$$K = {}^*\sum_{i=r}^{0} a_i\,\alpha_i$$

where a_i are integers such that $0 \leqslant a_i \leqslant \lambda_i$, and L in the form

$$L = {}^*\sum_{i=-1}^{-\infty} a_i\,\alpha_i$$

where again $0 \leqslant a_i \leqslant \lambda_i$, but for infinitely many i, $a_i < \lambda_i$. To show that an integer K is representable as described above we will use complete induction. $1 = \alpha_0$, hence 1 is representable. Assume that all integers $T < \alpha_i$ are representable in the form

$$T = {}^*\sum_{r=i-1}^{0} b_r\,\alpha_r.$$

Let $\alpha_i \leqslant K < \alpha_{i+1}$. Consider the intervals

$$\alpha_i \leqslant x < 2\alpha_i,\ 2\alpha_i \leqslant x < 3\alpha_i,\ 3\alpha_i \leqslant x < 4\alpha_i, \ldots,$$
$$(\lambda_i - 1)\alpha_i \leqslant x < \lambda_i\,\alpha_i,\ \lambda_i\,\alpha_i \leqslant x < \alpha_{i+1}.$$

Since these intervals cover the range $\alpha_i \leqslant x < \alpha_{i+1}$, K must lie in one of these intervals. If $r\alpha_i \leqslant K < (r+1)\alpha_i$, $1 \leqslant r < \lambda_i$ then $0 \leqslant K - r\alpha_i < \alpha_i$. By hypothesis

$$K - r\alpha_i = {}^*\sum_{t=i-1}^{0} c_t\,\alpha_t.$$

Hence

$$K = {}^*\sum_{t=i}^{0} c_t\,\alpha_t$$

where $1 \leqslant c_i = r \leqslant \lambda_i - 1$. If K lies in the last interval, i.e. $\lambda_i\,\alpha_i \leqslant K < \alpha_{i+1}$, then

$$0 \leqslant K - \lambda_i\,\alpha_i < \alpha_{i+1} - \lambda_i\,\alpha_i$$

$$= 1 + {}^*\sum_{t=i}^{0} \lambda_t\,\alpha_t - \lambda_i\,\alpha_i = 1 + {}^*\sum_{t=i-1}^{0} \lambda_t\,\alpha_t = \alpha_i.$$

Since $K - \lambda_i\,\alpha_i < \alpha_i$, by our induction hypothesis

$$K - \lambda_i\,\alpha_i = {}^*\sum_{t=i-1}^{0} c_t\,\alpha_t.$$

Hence

$$K = {}^*\sum_{t=i}^{0} c_t\,\alpha_t$$

where $c_i = \lambda_i$. We now come to the representation of L where $0 \leqslant L < 1$. We define a sequence $a_{-1}, a_{-2}, a_{-3}, \ldots, a_{-n}, \ldots$ as follows. First, determine the positive (or zero) integer a_{-1} by the condition that a_{-1} is the maximal admissible integer such that $a_{-1}\,\alpha_{-1} \leqslant L$. The maximal admissible integer a_{-1} is defined as follows. If $\lambda_{-1}\,\alpha_{-1} \leqslant L$ take $a_{-1} = \lambda_{-1}$. If $\lambda_{-1}\,\alpha_{-1} > L$ we choose a_{-1} to be the unique positive integer satisfying both the inequalities $a_{-1}\,\alpha_{-1} \leqslant L$ and $(a_{-1}+1)\alpha_{-1} > L$. In any case, $0 \leqslant a_{-1} \leqslant \lambda_{-1}$. We then choose a_{-2} as the maximal admissible integer satisfying $a_{-1}\,\alpha_{-1} + a_{-2}\,\alpha_{-2} \leqslant L$. In general, a_{-i} is the maximal admissible integer satisfying

$$a_{-1}\,\alpha_{-1} + a_{-2}\,\alpha_{-2} + \ldots + a_{-i}\,\alpha_{-i} \leqslant L.$$

Since for all i,

$${}^*\sum_{t=-1}^{i} a_t\,\alpha_t \leqslant L$$

the series

$${}^*\sum_{t=-1}^{-\infty} a_t\,\alpha_t$$

is convergent and

$$L \geqslant {}^*\sum_{i=-1}^{-\infty} a_i \alpha_i.$$

We next show that for infinitely many i, $a_i < \lambda_i$. In the first place, not all $a_i = \lambda_i$ since

$${}^*\sum_{i=-1}^{-\infty} \lambda_i \alpha_i = \alpha_0 = 1 > L.$$

Let r be the first index such that $a_{-r} < \lambda_{-r}$. By the maximality condition on a_{-r} we have

$${}^*\sum_{i=-1}^{-r} a_i \alpha_i \leqslant L < {}^*\sum_{i=-1}^{-r} a_i \alpha_i + \alpha_{-r}.$$

We now show that there is an $r_1 > r$ such that

$$a_{-r_1} < \lambda_{-r_1} .$$

If this were not so, then for all $i > r$, $a_{-i} = \lambda_{-i}$, from which it follows that

$$L \geqslant {}^*\sum_{i=-1}^{-\infty} a_i \alpha_i = {}^*\sum_{i=-1}^{-r} a_i \alpha_i + {}^*\sum_{i=-r-1}^{-\infty} \lambda_i \alpha_i = {}^*\sum_{i=-1}^{-r} a_i \alpha_i + \alpha_{-r} > L,$$

a contradiction. By continuation of this process one can obtain an infinite sequence

$$r < r_1 < r_2 < r_3 \ldots < r_n < \ldots$$

having the property that

$$a_{-r_i} \neq \lambda_{-r_i}$$

Furthermore, for any such r_i,

$${}^*\sum_{i=-1}^{-r_i} a_i \alpha_i \leqslant L < {}^*\sum_{i=-1}^{-r_i} a_i \alpha_i + \alpha_{-r_i}.$$

Since $\alpha_{-r_i} \to 0$ as $i \to \infty$ and since

$${}^*\sum_{i=-1}^{-\infty} a_i \alpha_i$$

is a series of positive terms, we have

$$L = {}^*\sum_{i=-1}^{-\infty} a_i \alpha_i.$$

We now show that the representation we have obtained for the number N is unique. If not, suppose there are two representations of N, namely

$$N = {}^*\sum^{-\infty} a_j \alpha_j = {}^*\sum^{-\infty} b_j \alpha_j$$

where $a_r \neq 0$, $b_s \neq 0$, $0 \leqslant a_j \leqslant \lambda_j$, $0 \leqslant b_j \leqslant \lambda_j$ and for infinitely many i, $a_i < \lambda_i$, for infinitely many k, $b_k < \lambda_k$. We will show that $r = s$ and $a_k = b_k$ for $r \geqslant k > -\infty$ If $r \neq s$, suppose $r > s$; then

$$*\sum_{j=r}^{-\infty} a_j \alpha_j \geqslant \alpha_r = *\sum_{j=r-1}^{-\infty} \lambda_j \alpha_j > *\sum_{j=s}^{r} b_j \alpha_j,$$

a contradiction. Hence $r = s$. If now $a_r > b_r \geqslant 1$, then

$$*\sum_{j=r}^{-\infty} a_j \alpha_j \geqslant a_r \alpha_r = (a_r - 1)\alpha_r + \alpha_r$$

$$= (a_r - 1)\alpha_r + *\sum_{i=r-1}^{-\infty} \lambda_i \alpha_i > *\sum_{i=r}^{-\infty} b_i \alpha_i,$$

a contradiction Hence $a_r = b_r$ and

$$*\sum_{j=r-1}^{-\infty} a_j \alpha_j = *\sum_{j=r-1}^{-\infty} b_j \alpha_j.$$

As before, $a_{r-1} = b_{r-1}$ and, in general, by inudction $a_i = b_i$ for $r \geqslant i > -\infty$.

Proof of necessity. Assume that the condition

$$\alpha_{r+1} = *\sum_{i=r}^{-\infty} \lambda_i \alpha_i, \qquad -\infty < r < \infty$$

is false. The following possibilities arise:

Case 1. There is at least one r for which

$$\alpha_{r+1} > *\sum_{i=r}^{-\infty} \lambda_i \alpha_i.$$

Case 2. For all r, it is false that

$$\alpha_{r+1} > *\sum_{i=r}^{-\infty} \lambda_i \alpha_i.$$

Case (2) breaks up into 2 sub-cases namely:

Case $(2)_1$.

$$*\sum_{i=r}^{-\infty} \lambda_i \alpha_i$$

is divergent for $-\infty < r < \infty$ and,

Case $(2)_2$.

$$\alpha_{r+1} \leqslant *\sum \lambda_i \alpha_i$$

for $-\infty < r < \infty$ and for at least one k,

$$\alpha_{r+1} < {}^*\sum_{i=r}^{-\infty} \lambda_i \alpha_i.$$

In case (1) let T be a number such that

$$\alpha_{r+1} > T > {}^*\sum_{i=r}^{-\infty} \lambda_i \alpha_i.$$

By definition, T has no representation. In case $(2)_1$, i.e. when

$${}^*\sum_{i=r}^{-\infty} \lambda_i \alpha_i$$

is divergent, define numbers $a_r, a_{r-1}, a_{r-2}, \ldots$ by saying that a_r is the maximal admissible integer such that $a_r \alpha_r \leqslant \alpha_{r+1}$; a_{r-1} is the maximal admissible integer such that $a_r \alpha_r + a_{r-1} \alpha_{r-1} \leqslant \alpha_{r+1}$ etc. Hence, by definition

$${}^*\sum_{i=r}^{-\infty} a_i \alpha_i$$

is convergent and

$$\alpha_{r+1} \geqslant {}^*\sum_{i=r}^{-\infty} a_i \alpha_i.$$

Since

$${}^*\sum_{i=r}^{-\infty} \lambda_i \alpha_i$$

is divergent, infinitely many $a_i < \lambda_i$. Let $i_1 < i_2 < i_3 < i_4 \ldots$ be an infinite sequence of integers such that

$$a_{-i_k} < \lambda_{-i_k}.$$

As before, we have

$${}^*\sum_{i=r}^{-i_k} a_i \alpha_i \leqslant \alpha_{r+1} < {}^*\sum_{i=r}^{-i_k} a_i \alpha_i + \alpha_{-i_k}.$$

Hence

$$0 \leqslant \alpha_{r+1} - {}^*\sum_{i=r}^{-i_k} a_i \alpha_i < \alpha_{-i_k}.$$

Since $\alpha_{-i_k} \to 0$ as $i_k \to \infty$, we have

$$\alpha_{r+1} = {}^*\sum_{i=r}^{-\infty} a_i \alpha_i.$$

We thus have two representations for α_{r+1}, namely

$$\alpha_{r+1} = {}^*\sum_{i=r}^{-\infty} a_i \alpha_i$$

and

$$\alpha_{r+1} = 1\alpha_{r+1} + 0\alpha_r + 0\alpha_{r-1} + \ldots.$$

In the case $(2)_2$ we are given that

$$\alpha_{k+1} < {}^*\sum_{i=k}^{-\infty} \lambda_i \alpha_i, \quad \alpha_{r+1} \leqslant {}^*\sum_{i=r}^{-\infty} \lambda_i \alpha_i, \qquad -\infty < r < \infty.$$

As before, define $a_k, a_{k-1}, a_{k-2}, \ldots$ by saying that a_k is the maximal admissible integer such that $a_k \alpha_k \leqslant \alpha_{k+1}$; a_{k-1} is the maximal admissible integer such that $a_k \alpha_k + a_{k-1} \alpha_{k-1} \leqslant \alpha_{l+1}$ etc. Then

$$\alpha_{k+1} \geqslant {}^*\sum_{i=k}^{-\infty} a_i \alpha_i.$$

At least one $a_i \neq \lambda_i$ and hence, $a_i < \lambda_i$, since

$$\alpha_{k+1} < {}^*\sum_{i=l}^{-\infty} \lambda_i \alpha_i.$$

If there were only a finite number of i for which $a_i < \lambda_i$ let $n_0 + 1$ be the last index for which $a_i < \lambda_i$. That is, we assume that

$$a_{n_0+1} < \lambda_{n_0+1}$$

but that $a_i = \lambda_i$ for all $i \leqslant n_0$ (remembering that in the series ${}^*\sum a_i \alpha_i$ the i run in decreasing order of magnitude). Then

$$\alpha_{k+1} \geqslant {}^*\sum_{i=k}^{n_0+1} a_i \alpha_i + {}^*\sum_{i=n_0}^{-\infty} \lambda_i \alpha_i \geqslant {}^*\sum_{i=k}^{n_0+1} a_i \alpha_i + \alpha_{n_0+1}.$$

But since a_{n_0+1} is a maximal admissible number and

$$a_{n_0+1} < \lambda_{n_0+1},$$

we have

$${}^*\sum_{i=k}^{n_0+1} a_i \alpha_i \leqslant \alpha_{k+1} < {}^*\sum_{i=k}^{n_0+1} a_i \alpha_i + \alpha_{n_0+1},$$

a contradiction. Therefore, infinitely many $a_i < \lambda_i$. As before, α_{k+1} has at least two representations. We have shown that when the condition

$$\alpha_{r+1} = {}^*\sum_{i=r}^{-\infty} \lambda_i \alpha_i, \qquad -\infty < r < \infty,$$

is violated, either there exist positive real numbers with no representations or there exist positive real numbers with at least two representations. Hence, we cannot have uniqueness of representation for every positive real number if we violate the fundamental condition: i.e. the condition is necessary.

3. Some Subsidiary Theorems

Theorem 2. *The positive integers λ_i, $-\infty < i < \infty$, may be chosen arbitrarily and these uniquely determine the α_i and the β_i.*

Proof. It can easily be shown by induction that

$$\alpha_0 = 1, \ \alpha_n = (1 + \lambda_0)(1 + \lambda_1) \dots (1 + \lambda_{n-1})$$

and that

$$\beta_n = (1 + \lambda_{-1})(1 + \lambda_{-2}) \dots (1 + \lambda_{-n})$$

THEOREM 3. *The numbers* α_n *and* β_n *are integers such that for* $i \leqslant j$, $\alpha_i \mid \alpha_j$ *and* $\beta_i \mid \beta_j$. ($x \mid y$ means x is a divisor of y.)

Proof. The result follows immediately from the formulae of Theorem 2.

THEOREM 4. *The numbers* α_i *and* β_i *may be taken as arbitrary sequences of integers subject only to the conditions* $\alpha_0 = 1$, $\alpha_i \mid \alpha_{i+1}$, *and* $\beta_i \mid \beta_{i+1}$ *for all* i.

Proof. If we choose

$$\lambda_n = \frac{\alpha_{n+1}}{\alpha_n} - 1,$$

and

$$\lambda_{-n} = \frac{\mu_n}{\beta_{n-1}} - 1, \qquad n \geqslant 2, \ \lambda_{-1} = \beta_1 - 1.$$

it is easily verified that the fundamental condition is satisfied.

The result of the Theorem 4 shows that Cantor in [1] actually obtained the most general representation (for the fractional part of the number at least).

THEOREM 5. *A necessary and sufficient condition that every positive rational number have a finite expansion, is that for every integer* N, *there exists an* n *such that* $N \mid \beta_n$.

We note here that a proof of half of this theorem was obtained by Cantor in [1].

Proof of necessity. Suppose there is an integer N such that N does not divide β_n for $n = 1, 2, 3, \dots$. We will show that $1/N$ cannot have a finite expansion. For suppose that

$$\frac{1}{N} = \sum_{i=1}^{l} \frac{a_i}{\beta_i} = \frac{u}{\beta_k},$$

since $\beta_1 \mid \beta_2 \mid \beta_3 \dots \mid \beta_k$. Then $\beta_k = Nu$, that is, $N \mid \beta_k$ which is a contradiction.

Proof of sufficiency. Suppose that for each N, there is an n such that $N \mid \beta_n$. We need only consider the representation of rational numbers a/b such that $0 < a/b < 1$. Suppose $b \mid \beta_1$. Then $a/b = r/\beta_1$. If $r \leqslant \lambda_{-1}$, r/β_1 is the required representation. If $r > \lambda_{-1}$ then

$$\frac{a}{b} = \frac{r}{\beta_1} \geqslant \frac{\lambda_{-1} + 1}{\beta_1} = \alpha_{-1} + \alpha_{-1}\,\lambda_{-1}$$

$$= \alpha_{-1}\,\lambda_{-1} + {}^*\sum_{i=-2}^{-\infty} \lambda_i\,\alpha_i = {}^*\sum_{i=-1}^{-\infty} \lambda_i\,\alpha_i = \alpha_0 = 1,$$

a contradiction. Hence every number a/b such that $0 < a/b < 1$ and such that $b \mid \beta_1$ has a representation by one term. Suppose now that any fraction whose denominator divides β_r has a finite representation stopping at the term in β_r and let c/d be a fraction such that $0 < c/d < 1$ and such that $d \mid \beta_{r+1}$. Then $c/d = k/\beta_{r+1}$. If $k \leqslant \lambda_{-r-1}$, k/β_{r+1} is the required representation of c/d by one term. If $k > \lambda_{-r-1}$, divide k by $\lambda_{-r-1} + 1$ obtaining $k = q\,(\lambda_{-r-1} + 1) + s$ where $0 \leqslant s \leqslant \lambda_{-r-1}$. Then

$$\frac{c}{d} = \frac{k}{\beta_{r+1}} = \frac{q(\lambda_{-r-1} + 1) + s}{\beta_{r+1}} = \frac{q}{\beta_r} + \frac{s}{\beta_{r+1}}.$$

By hypothesis

$$\frac{q}{\beta_r} = \frac{c_1}{\beta_1} + \frac{c_2}{\beta_2} + \dots + \frac{c_r}{\beta_r}$$

where $0 \leqslant c_i \leqslant \lambda_{-i}$, and hence,

$$\frac{c}{d} = \frac{c_1}{\beta_1} + \frac{c_2}{\beta_2} + \dots + \frac{c_r}{\beta_r} + \frac{s}{\beta_{r+1}}.$$

THEOREM 6. *If the system of representation defined by the sequences* $\{\alpha_n\}$ *and* $\{\beta_n\}$ *represents all rational numbers by finite expansions, then* $\{\lambda_{-n}\}$ *is an unbounded sequence.*

Proof. If $\lambda_{-n} \leqslant M$ for all n, take p a prime number such that $p > M + 1$. For all n,

$$\beta_n = (1 + \lambda_{-1})(1 + \lambda_{-2}) \dots (1 + \lambda_{-n}).$$

Since all the numbers $1 + \lambda_{-1}, 1 + \lambda_{-2}, \dots, 1 + \lambda_{-n}$ are less than or equal to $1 + M$, all the prime factors of β_n are less than or equal to $1 + M$. Hence p is not a divisor of any β_n. By Theorem 5, not all rational numbers have a finite expansion, a contradiction. Hence, $\{\lambda_{-n}\}$ is unbounded.

THEOREM 7. *Suppose that* $\beta_{n+1} \geqslant \beta_n^n$ *for all* $n \geqslant n_0$. *If a number* T *be represented by a bounded representation, that is,*

$$T = {}^*\sum_{i=r}^{-\infty} a_i\,\alpha_i,$$

and $a_i \leqslant A$ *for all* i *but infinitely many* $a_i \neq 0$, *then* T *is a transcendental number.*

Proof. Put

$$T_N = {}^*\sum_{i=r}^{-N} a_i \alpha_i.$$

Since infinitely many $a_i \neq 0$, $T_N < T$ for all N. Also,

$$T_N = {}^*\sum_{i=r}^{-N} a_i \alpha_i = \frac{K}{\beta_N}.$$

Then,

$$\begin{aligned} 0 < T - \frac{K}{\beta_N} = T - T_N &= {}^*\sum_{i=-N-1}^{-\infty} a_i \alpha_i = a_{-N-1}\alpha_{-N-1} + {}^*\sum_{i=-N-2}^{-\infty} a_i \alpha_i \\ &< a_{-N-1}\alpha_{-N-1} + {}^*\sum_{i=-N-2}^{-\infty} \lambda_i \alpha_i = a_{-N-1}\alpha_{-N-1} + \alpha_{-N-1} \\ &= \alpha_{-N-1}(a_{-N-1} + 1) < \frac{A+1}{\beta_{N+1}} < \frac{A+1}{\beta_N{}^N} \quad \text{for all } N \geqslant n_0. \end{aligned}$$

By Liouville's theorem, the inequalities

$$0 < T - \frac{K}{\beta_N} < \frac{A+1}{\beta_N{}^N},$$

for all $N \geqslant n_0$, implies that T is transcendental.

COROLLARY. *Let* $\beta_{n+1} \geqslant \beta_n{}^n$ *for all* $n \geqslant n_0$. *If a number* T *be represented by*

$$T = {}^*\sum_{i=n}^{-\infty} a_i \alpha_i$$

and $a_i \geqslant \lambda_i - A$, *all* i, *then* T *is transcendental. For if* $T = K + L$ *where* K *is integral and* $0 \leqslant L \leqslant 1$, *then, by Theorem 7,* $1 - L$ *is transcendental and hence so is* T.

REFERENCES

1. G. CANTOR, *Über die Einfachen Zahlensysteme*, Schlömilch's Zeitschrift, vol. 14 (1869), 121.

2. G. FABER, Math. Ann., vol. 60 (1905), 196.

TRANSACTIONS OF

THE ROYAL SOCIETY OF CANADA

SECTION III

CHEMICAL, MATHEMATICAL, AND PHYSICAL SCIENCES

THIRD SERIES—VOLUME XLVII—SECTION III

JUNE, 1953

40873

OTTAWA
THE ROYAL SOCIETY OF CANADA
1953

TABLE OF CONTENTS

Convex Sets in Linear Topological Spaces. By ISRAEL HALPERIN, F.R.S.C. 1

Bessel Expansions of the Confluent Hypergeometric Functions. By T. E. HULL, C. A. SWANSON, and D. A. TRUMPLER . . 7

The Ergodic Theorem for Banach Spaces with Convex-Compactness. By VALENTINA KROTKOV and ISRAEL HALPERIN, F.R.S.C. 17

A Problem in Combinatorial Analysis. By N. S. MENDELSOHN 21

Note on a Combinatorial Formula of Mendelsohn. By LEO MOSER 27

Some Remarks on Laplace's Method. By P. G. ROONEY . . . 29

A Remark on Curves of Order n *in* n*-space.* By P. SCHERK, F.R.S.C. 35

The Application of Fourier Transforms in Physical Problems. By W. H. WATSON, F.R.S.C. 37

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLVII : SERIES III : JUNE, 1953
SECTION THREE

Convex Sets in Linear Topological Spaces

ISRAEL HALPERIN, F.R.S.C.

1. INTRODUCTION

IN this paper we shall prove some theorems concerning convex sets in linear topological spaces.

In §2 we shall show that if K and K_1 are non-empty, disjoint convex sets in a real-linear topological space and if at least one of them has inner points then there is a hyperplane separating them. This was proved by M. Eidelheit [3] for the more special case that the space is linear and normed and both K and K_1 have inner points.

From this result we deduce that if K is a convex set with inner points then through every boundary point of K there passes a supporting hyperplane; more generally if R is a linear set containing no inner points of K then there is a hyperplane H containing R such that K lies on one side of H. This was proved for linear normed spaces by S. Mazur [4], the first half of this result having been obtained previously by G. Ascoli [1] for separable, linear, normed spaces.

Aside from greater generality, our proofs are more elementary than those of Ascoli, Mazur, and Eidelheit and do not use Minkowski functionals or the Hahn-Banach extension theorem. (The simplified proofs for Eidelheit's theorem due to Kakutani [5] and Botts [2] require the theorem of Ascoli-Mazur.)

In §3, we consider linear topological spaces with either the real or complex numbers as scalars. We repeat the argument of Ascoli and Mazur to obtain the theorems (proved by them for the real-linear normed case) that every closed convex set is weakly closed and that if $x_n \to x$ weakly then x is in the closure of the set of all averages $c_1x_1 + \ldots + c_mx_m$ with $m = 1, 2, \ldots$, all $c_n \geqslant 0$ and the sum of the c_n equal to 1; however, for this result we require the stronger restriction on the space that every open set containing the origin should contain a convex set having the origin as inner point.

2. SEPARATING HYPERPLANES

Let E be a linear space topologized by a family of open sets such that: the set-union of an arbitrary collection (possibly empty) of open sets is an open set, the set-intersection of two open sets is an open set, and each point of E is the intersection of all open sets containing it. We shall suppose (i) $x + y$ is continuous in x for fixed y and (ii) cx is continuous in the pair c, x. It follows that whenever 0 is in the closure of a set of x then 0 is also in the closure of the cx where c may depend on x providing that $K = \sup |c|$ is finite; indeed, for arbitrary integer p each cx may be expressed as $m2^{-p}x + r_p x$ with $0 \leqslant r_p < 2^{-p}$ and m one of the finite set of integers numerically less than $2^p(2 + K)$, and hence by choosing p sufficiently large it can be seen that in every open set containing 0 there is at least one cx.

If S_1, S_2 are subsets (possibly single points) of E then $S_1 + S_2$ shall denote the set of all $x_1 + x_2$ with x_1 in S_1 and x_2 in S_2, and cS the set of all cx with x in S; $J(S_1, S_2)$, the *join* of S_1 and S_2, shall denote the set of all $cx_1 + (1 - c)x_2$ with x_1 in S_1, x_2 in S_2, and $0 \leqslant c \leqslant 1$. Our conditions on E imply that $x + N$ and cN are open sets for fixed x and $c \neq 0$ respectively if N is an open set.

A set K is called convex if $K = J(K, K)$. It is easily verified that $J(S_1, S_2)$ is convex if S_1, S_2 are both convex and that the set union of an increasing family of convex sets is again convex.

Our conditions on E ensure that if K is convex then its closure $\bar{K}$ is also convex; for if $K \subset K_1$ for some convex $K_1 \subset \bar{K}$ and if x is in $\bar{K}$ then $J(x, K_1)$ is again convex and $K \subset K_1 \subset J(x, K_1) \subset \bar{K}$ and it follows that K is contained in a maximal such K_1 which must then coincide with $\bar{K}$.

For the rest of this section we shall suppose that the scalars are the real numbers.

$l(x, y)$ shall denote the set of all $tx + (1 - t)y$ with $-\infty < t < \infty$; $l(x)$ shall mean $l(0, x)$. A set H is called linear if H contains $l(x, y)$ for all x, y in H; H is called a *hyperplane* if it is closed, linear, and different from E and the only linear sets which contain H are H and E.

Suppose H is a hyperplane. Choose any x_0 in H, any y_0 not in H and set $u = y_0 - x_0$. Then the linear set $H + l(u)$ must coincide with E and hence every x in E can be represented in the form $h + tu$ with h in H. Clearly h and t are determined uniquely by x. Let 0 have the representation $h_0 + t_0 u$ and for every $x = h + tu$ set $f(x) = t - t_0$. Then $f(x)$ is a real-valued linear functional on E and our conditions on E together with the closure of H ensure that $f(x)$ is continuous. (If $f(x)$ were not continuous we could deduce that for some $\epsilon > 0$ and some h_1, t_1 and some set of h, t, $|t - t_1| > \epsilon$ but $h_1 + t_1u$ is in the closure of $h + tu$; hence 0 is in the closure of the set of $h - h_1 + (t - t_1)u$; then, since $|t - t_1|^{-1} < \epsilon^{-1}$, 0 is in the closure of the set of $(t - t_1)^{-1}(h - h_1) + y_0 - x_0$ and finally y_0 is in the closure of H.) H consists precisely of the x for which $f(x) = -t_0$. Each of the sets of x for which $f(x) \leqslant -t_0$, and $f(x) \geqslant -t_0$ respectively, is called *a side of H* and also a *half-space*. It is easily verified that the two sides of H are uniquely determined by H and that if y is an inner point of a set S lying on one side of H then y is not in H.

Conversely, if H is a closed linear set with $H \neq E = H + l(u)$ for some u then it is easily verified that H is a hyperplane. Also, if $f(x)$ is any real-valued continuous linear functional not identically zero and c is a real number then the set of x for which $f(x) = c$ is a hyperplane.

We shall say that two sets S_1 and S_2 are separated by the hyperplane H if they lie on different sides of H, equivalently if, for some continuous real-valued linear functional $f(x)$ and some c, $f(x) = c$ if and only if x is in H, $f(x) \leqslant c$ for all x in S_1 and $f(x) \geqslant c$ for all x in S_2.

A hyperplane H is called a *supporting* hyperplane of a set S if S lies on one side of H and H contains at least one boundary point of S.

LEMMA. *If C is a closed set with inner points and its boundary H is a non-empty linear set then H is a hyperplane and C is one of its sides.*

Proof. (i) By a translation, if necessary, we may suppose that the origin 0 is in H.

(ii) Since C is closed and its boundary is a linear set it follows that C is convex. Hence if u is in C then tu is in C for all $t \geqslant 0$ and $h + tu = (1 + t)\,((1 + t)^{-1}h + t(1 + t)^{-1}u)$ is in C for all h in H and $t \geqslant 0$.

(iii) If u and $-u$ are both in C then u is in H. For otherwise there would be an open set N containing u and contained in C. It would follow that $J(N, -u)$ is contained in C and hence $\frac{1}{2}(N + (-u))$, which is an open set containing the origin, is contained in C, contradicting the fact that the origin is a boundary point of C.

(iv) If u_0 is a fixed point in C but not in H then C consists precisely of all $h + tu_0$ with h in H and $t \geqslant 0$. Indeed, if v in C were not of this form it would follow that for every $t \geqslant 0$, $v - tu_0$ is not in H. Since v would be an inner point of C it would follow that for every $t > 0$, $v - tu_0$ is in C, $t^{-1}(v - tu_0) = t^{-1}v - u_0$ is in C. When $t \to \infty$ we obtain $-u_0$ is in C, assumed closed. By (iii) it would follow that u_0 is in H contradicting our assumption.

(v) With the u_0 of (iv), $H + l(u_0) = E$ since it is a linear set having inner points. Since H is closed this implies the lemma.

Theorem 1. *Suppose that K and K_1 are non-empty disjoint convex sets and that K is open. Then there is a continuous linear $f(x)$ and a c such that $f(x) < c$ for all x in K and $f(x) \geq c$ for all x in K_1, so that the hyperplane $f(x) = c$ separates K and K_1.*

Proof. (i) Since the union of an increasing family of convex sets, each disjoint from K, is also a convex set disjoint from K we may suppose that K_1 is not contained in any other convex set disjoint from K. Then K_1 will be closed since the closure of K_1 is convex and disjoint from the open set K.

(ii) If u is not in K_1 then $J(u, K_1) \neq K_1$. But the set $J(u, K_1)$ is convex and contains K_1 so it must intersect K. Thus $cu + (1 - c)x_1$ must be in K for some x_1 in K_1 and some $0 < c \leq 1$.

(iii) If x_1 is in K_1 and u is arbitrary then at least one of the sets: all $x_1 + tu$, $t \geq 0$ *or* all $x_1 - tu$, $t \geq 0$ is contained in K_1. Otherwise there would exist u_1 and u_2, neither in K_1 and with x_1 in $J(u_1, u_2)$. Then by (ii) there would be y_1, z_1 with z_1 in K_1, y_1 in K and y_1 contained in $J(u_1, z_1)$; similarly for u_2, y_2, z_2. Then $J(y_1, y_2)$ would intersect the triangle of x_1, z_1 and z_2, that is, $J(x_1, J(z_1, z_2))$, in a point w which would be in both K and K_1, a contradiction.

(iv) If x_1 is in K_1 but u is not in K_1 then for every $t > 0$, $x_1 + t(x_1 - u)$ is an inner point of K_1. For there is an open set N containing u and disjoint from K_1. Then $x_1 + t(x_1 - y)$ is in K_1 for all $t \geq 0$ and all y in N, and for fixed $t > 0$, $x_1 + t(x_1 + (-1)N)$ is an open set containing $x_1 + t(x_1 - u)$.

(v) Since K_1 is a convex set it now follows that the boundary of K_1 is a linear set and that K_1 is a closed set with inner points. The theorem then follows from the previous lemma.

Corollary 1. *If K and K_1 are non-empty convex sets and K has inner points, then K and K_1 can be separated by a hyperplane whenever K_1 is contained in the closure of a convex set K' such that K' contains no inner point of K, in particular if K_1 contains inner points but K and K_1 have no common inner points.*

Proof. This follows from the fact that if a convex set has inner points it is contained in the closure of the convex set of its inner points.

Corollary 2. *If x is a boundary point of a convex set K possessing inner points then there exists a supporting hyperplane of K which contains x.*

Proof. This follows from Corollary 1 if x is taken as K_1.

Theorem 2. *Suppose that K is a convex set with inner points and* *that R is a non-empty linear set containing no inner points of K. Then R is contained in a hyperplane H such that K lies on one side of H.*

Proof. Replacing K by the convex set of its inner points, we may assume that K is an open set and that K and R are disjoint.

Set $J = J(R, K)$. Then J is a convex set with inner points. Moreover no inner point of J is in R. Indeed, if x in R were an inner point of J we could choose any x_0 in K and deduce that for some $t > 0$, $x + t(x - x_0) = y_1$ is in J but not in R and x is in $J(x_0, y_1)$. Then y_1 would be in $J(x_1, y)$ for some x_1 in K and y in R. This would imply that $l(x, y)$, all in R, intersects $J(x_0, x_1)$ all in K, contradicting the assumption that K and R are disjoint.

Corollary 1 to Theorem 1 now implies that J and R can be separated by a hyperplane H. Since R is contained in J, R must lie on both sides of H and hence is contained in H.

THEOREM 3. *If E is such that every open set containing 0 contains a convex subset with 0 as an inner point, then if K is a non-empty closed convex set and y is not in K there is an open set N containing y such that K and N can be separated by a hyperplane H. (Necessarily, y is not in H.)*

Proof. By our assumptions on E there will be an open set N containing y with N contained in a convex set K_1 such that K and K_1 are disjoint. Since K_1 has inner points it follows from Corollary 1 to Theorem 1 that K and K_1, *a fortiori* K and N, can be separated by a hyperplane.

COROLLARY. *For E as in Theorem 3, every closed convex set different from E is the intersection of the half-spaces which contain it.*

3. WEAK CONVERGENCE

If E is a linear space with a topology as specified in §2 we shall say that a sequence x_n in E converges *weakly* to x if $f(x_n)$ converges to $f(x)$ for every continuous linear $f(x)$ on E which is real or complex valued according as the scalars for E are the real or complex numbers. A set Z contained in E is called *weakly* closed if x is in Z whenever a sequence x_n in Z converges weakly to x.

If the scalars are the complex numbers then E may be regarded also as a linear space E' with the real numbers as scalars and with the same open sets as E. If $f(x)$ is complex-valued and linear and continuous on E, then clearly $\phi(x) = \mathcal{R} f(x)$ is real-valued, linear and continuous on E' and conversely for every real-valued, linear, continuous $\phi(x)$ on E', $f(x) = \phi(x) - i\phi(ix)$ is complex-valued, linear and continuous on E.

THEOREM 4. *If every open set containing the origin has a convex subset with the origin as inner point, then:*

(i) *Every closed convex set is weakly closed.*

(ii) *If x_n converges weakly to x then x is in the closure of the set of all averages $c_1x_1 + \ldots + c_mx_m$ with $m = 1, 2, \ldots$, all $c_n \geqslant 0$ and the sum of the c_n equal to 1. In particular, if E satisfies the first countability axiom of Hausdorff there will be a sequence of such averages which converge to x in the sense of the open-set topology.*

Proof. We may suppose that the scalars are the real numbers since x_n converges weakly to x for complex scalars if and only if this is true for real scalars.

If now K is a closed convex set and x_n in K converges weakly to y and y is not in K then Theorem 3 implies that there is a hyperplane H separating K and y with y not in H. Then there is a continuous real-valued linear $f(x)$ and a real number c such that $f(x) \leqslant c$ for all x in K but $f(y) > c$. This contradicts the fact that $f(x_n)$ must converge to $f(y)$.

(After this note was accepted for publication the writer learned that an elementary proof of the Mazur-Ascoli theorem had been found earlier by M. H. Stone and others, see V. L. Klee, Jr., *Convex sets in linear spaces*, Duke Math. J., vol. 18 (1951), pages 443–66, 875–83 for references.)

REFERENCES

1. G. ASCOLI, *Sugli spazi lineari metrici e le loro varietà lineari*, Annali di Matematica, vol. 10 (1932), 33–81, 203–32.
2. T. BOTTS, *On convex sets in linear normed spaces*, Bull. Amer. Math. Soc., vol. 48 (1942), 150–2.
3. M. EIDELHEIT, *Zur Theorie der konvexen Mengen in linearen normierten Raumen*, Studia Math., vol. 6 (1936), 104–11.
4. S. MAZUR, *Über konvexe Mengen in linearen normierten Raumen*, Studia Math., vol. 4 (1933), 70–84.
5. S. KAKUTANI, *Ein Beweis des Satzes von M. Eidelheit uber konvexe Mengen*, Proc. Imp. Acad. Japan (Tokyo), vol. 13 (1937), 93–4.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLVII : SERIES III : JUNE, 1953
SECTION THREE

Bessel Expansions of the Confluent Hypergeometric Functions*

T. E. HULL, C. A. SWANSON, and D. A. TRUMPLER

Presented by R. D. JAMES, F.R S.C.

INTRODUCTION

An initial or boundary value problem associated with a differential equation can be reformulated in terms of an integral equation in a variety of ways. When the problem cannot be solved exactly, such an integral equation is usually the starting point of any successive approximation procedure used to generate the solution. It is also natural to begin with such a reformulation when investigating the asymptotic behaviour of the solution, or, in the case of boundary value problems, when investigating the convergence of an eigenfunction expansion or the asymptotic behaviour of eigenvalues and eigenfunctions.

We shall first discuss briefly the reformulation of a general initial value problem in terms of an integral equation. We then adapt the method of successive substitutions to a special case and obtain the multiplication theorems for the Bessel functions. An extension of the idea leads to expansions of the confluent hypergeometric functions in series of Bessel functions.

We shall include references to the other applications mentioned above and to other expansions of the confluent hypergeometric functions which have been found by different methods.

THE INTEGRAL EQUATION

Suppose that after dividing through by $p(z)$, the coefficients in the differential equation

$$\frac{d}{dz}\left(p(z)\frac{dy}{dz}\right) + q(z)\,y = r(z)\,y \tag{1}$$

---

*This investigation was carried out while two of the authors (C.A.S. and D.A.T.) held National Research Council Scholarships.

are analytic in the finite z-plane except perhaps for poles at the origin. Then, for $\epsilon \neq 0$,

$$(2)\quad y(z) = a\, v_1(z) + b\, v_2(z) + c \int_\epsilon^z \{v_1(x)\, v_2(z) - v_1(z)\, v_2(x)\}\, r(x)\, y(x)\, dx,$$

where v_1, v_2 are independent solutions of

$$(3)\qquad \frac{d}{dz}\left(p(z)\frac{dv}{dz}\right) + q(z)\, v = 0$$

and where a, b are arbitrary constants while

$$c = 1/p(x)\{v_1(x)\, v_2'(x) - v_2(x)\, v_1'(x)\}\,.$$

The transformation from (1) to (2) links the solutions of (1) with those of the associated equation (3).

The transformation can be generalized to apply to n-th order equations. For example, with suitable restrictions on the coefficients, the differential equation

$$(1')\qquad y^{(n)}(z) + \sum_{i=0}^{n-1} q_i(z)\, y^{(i)}(z) = \sum_{i=0}^{n-1} r_i(z)\, y^{(i)}(z)$$

can be transformed into

$$(2')\qquad y(z) = \sum_{i=1}^{n} a_i\, v_i(z) + \int_\epsilon^z K(z, x)\, y(x)\, dx,$$

where the v_i are independent solutions of

$$(3')\qquad v^{(n)}(z) + \sum_{i=0}^{n-1} q_i(z)\, v^{(i)}(z) = 0.$$

$K(z, x)$ can be found by the method of variation of parameters and it turns out to be

$$K(z, x) = \sum_{i=0}^{n-1} (-1)^i \left\{N(z, x)\, r_i(x)\right\}^{(i)}$$

here

$$\begin{vmatrix} v_1(x) & \ldots\ldots & v_n(x) \\ v_1'(x) & \ldots\ldots & v_n'(x) \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ v_1^{(n-2)}(x) & \ldots & v_n^{(n-2)}(x) \\ v_1^{(n-1)}(x) & \ldots & v_n^{(n-1)}(x) \end{vmatrix} \quad N(z, x) = \begin{vmatrix} v_1(x) & \ldots\ldots & v_n(x) \\ v_1'(x) & \ldots\ldots & v_n'(x) \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ v_1^{(n-2)}(x) & \ldots & v_n^{(n-2)}(x) \\ v_1(z) & \ldots\ldots & v_n(z) \end{vmatrix}$$

The transformation from (1) to (2) was first used by Liouville [10] in the special case when $p(z) = 1$, $q(z) = \rho^2$; the resulting integral equation was then used to study the asymptotic behaviour, for large ρ, of the eigenvalues and eigenfunctions determined by (1) after the imposition of certain boundary conditions. More general problems of the same nature have been considered by Birkhoff [2; 3].

Many papers have been devoted to a study of the asymptotic behaviour of solutions of differential equations. A majority of these papers have been based on a reformulation of the problem in terms of an integral equation and the pertinent theory has been developed chiefly during the past twenty years. Recent developments are given, for example, by Langer [14] and Cherry [7]. A transformation of an n-th order differential equation has been considered in detail by Miller [15].

It should be noted that the terms on the right side of (1) or (1′) need not be linear in y and its derivatives. If these terms are non-linear, but relatively small, the form (2) or (2′) is useful in finding successive approximations to the solution if the problem is an initial value problem. A trivial but interesting example of this fact arises in relativity theory where the path of light past the sun depends on the solution of [8]

$$\frac{d^2u}{d\phi^2} + u = 3mu^2, u(0) = \frac{1}{R}, u'(0) = 0$$

where m is small. (Other methods are more direct when only the so-called limit cycle solutions of a non-linear problem are required.)

MULTIPLICATION THEOREMS FOR BESSEL FUNCTIONS

In the above mentioned studies of asymptotic expansions, the usual procedure has been to choose the integral equation (2) or (2′) so that the integral contained in it is of small order compared to the remaining terms, as some parameter approaches infinity. The remaining terms on the right side are then *asymptotic* to the desired solution $y(z)$. Apparently Ikeda [11] was the first to use the method of successive substitutions in (2) to obtain complete and ***convergent*** expansions of the solution in special cases. His results are not entirely correct; we shall therefore now treat his most important example in order to illustrate the procedure.

The differential equation for $J_{\pm\nu}(\alpha z)$ (ν not an integer) can be written in the form (1) as

$$(4) \qquad \frac{d}{dz}\left(z\frac{dy}{dz}\right) + \left(z - \frac{\nu^2}{z}\right) y = z(1 - \alpha^2)\, y$$

In this case $v_1 = J_\nu(z)$ and $v_2 = J_{-\nu}(z)$, and (2) becomes

$$(5)\quad y(z) = a\, J_\nu(z) + b\, J_\nu(z) - \frac{\pi(1-\alpha^2)}{2 \sin \nu\pi} \int_\epsilon^z \{J_{-\nu}(x)\, J_{-\nu}(z) - J_\nu(z)\, J_{-\nu}(x)\}\, x\, y(x)\, dx.$$

Here and in later sections we need, for $i = 0, 1, 2, \ldots\ldots,$ the results*

$$(6a)\quad \frac{-\pi}{2 \sin \nu\pi} \int_\epsilon^z \{J_\nu(x)\, J_{-\nu}(z) - J_\nu(z)\, J_{-\nu}(x)\}\, x\{x^{i+1}\, J_{\nu+i}(x)\}\, dx = \frac{z^{i+1}}{2(i+1)} J_{\nu+i+1}(z) + A_i J_\nu(z) + B_i J_{-\nu}(z),$$

where

$$(6b)\quad A_i = \frac{-\pi\epsilon}{4(i+1)\sin \nu\pi}\left[J_{-\nu}(\epsilon)\frac{d}{d\epsilon}\{\epsilon^{i+1} J_{\nu+i+1}(\epsilon)\} - \{\epsilon^{i+1} J_{\nu+i+1}(\epsilon)\}\frac{d}{d\epsilon} J_{-\nu}(\epsilon)\right],$$

$$(6c)\quad B_i = \frac{-\pi\epsilon}{4(i+1)\sin \nu\pi}\left[-J_\nu(\epsilon)\frac{d}{d\epsilon}\{\epsilon^{i+1} J_{\nu+i+1}(\epsilon)\} + \{\epsilon^{i+1} J_{\nu+i+1}(\epsilon)\}\frac{d}{d\epsilon} J_\nu(\epsilon)\right],$$

and

$$(7a)\quad \frac{-\pi}{2 \sin \nu\pi} \int_\epsilon^z \{J_\nu(x)\, J_{-\nu}(z) - J_\nu(z)\, J_{-\nu}(x)\}\, x\{x^{i}\, J_{-\nu-i}(x)\}\, dx = \frac{-z^{i+1}}{2(i+1)} J_{-\nu-i-1}(z) + C_i J_\nu(z) + D_i J_{-\nu}(z),$$

where

$$(7b)\quad C_i = \frac{\pi\epsilon}{4(i+1)\sin \nu\pi}\left[J_{-\nu}(\epsilon)\frac{d}{d\epsilon}\{\epsilon^{i+1} J_{-\nu-i-1}(\epsilon)\} - \{\epsilon^{i+1} J_{-\nu-i-1}(\epsilon)\}\frac{d}{d\epsilon} J_{-\nu}(\epsilon)\right],$$

$$(7c)\quad D_i = \frac{\pi\epsilon}{4(i+1)\sin \nu\pi}\left[-J_\nu(\epsilon)\frac{d}{d\epsilon}\{\epsilon^{i+1} J_{-\nu-i-1}(\epsilon)\} + \{\epsilon^{i+1} J_{-\nu-i-1}(\epsilon)\}\frac{d}{d\epsilon} J_\nu(\epsilon)\right].$$

---

*To avoid ambiguity in these results we will assume the z-plane to be cut, say along the negative real axis. The expansions obtained later will nevertheless hold without any restriction on arg z.

These results can be established as follows: it is a straight-forward matter to show that the left side of (6a) satisfies

$$\frac{d}{dz}\left(z\frac{du}{dz}\right) + \left(z - \frac{\nu^2}{z}\right) u = z^{i+1} J_{\nu+i}(z).$$

The first term on the right of (6a) is a particular integral of this equation and the other terms are the complementary function. A_i and B_i are found by putting $z = \epsilon$ in (6a) and in the equation obtained by differentiating (6a). The proof of (7) is analogous.

For each a and b, the method of successive substitutions applied to (5) will generate a series which converges to the solution of (5) for any finite z, $\epsilon \neq 0$. Assuming for the moment that we can rearrange the various contributions to the terms in this series, it is clear from (6) and (7) that we can write the solution in the form

$$(8) \qquad y(z) = \sum_{i=0}^{\infty} p_i z^i J_{\nu+i}(z) + \sum_{i=0}^{\infty} q_i z^i J_{-\nu-i}(z),$$

where p_i and q_i are constant coefficients.

Proceeding formally, we now substitute (8) into (5); using (6) and (7) and equating coefficients of like terms, we then find that (5) is satisfied if p_i and q_i satisfy the recurrence relations

$$(9) \qquad p_{i+1} = \frac{1-\alpha^2}{2(i+1)} p_i, \quad q_{i+1} = -\frac{1-\alpha^2}{2(i+1)} q_i, \quad i = 0, 1, 2, \ldots .$$

The values of p_0 and q_0 depend on a and b and so ultimately on the particular solution of (4) which is being considered

For the solution to be $J_\nu(\alpha z)$, which is finite at the origin when $R(\nu) > 0$, we must take $q_0 = 0$ in (8). Equating coefficients of z^ν on both sides of (8), we obtain the value of p_0. From (9) the solution then formally reduces to

$$(10) \qquad J_\nu(\alpha z) = \alpha^\nu \sum_{i=0}^{\infty} \frac{(1-\alpha^2)^i}{i!} \left(\frac{z}{2}\right)^i J_{\nu+i}(z)$$

which we shall now show holds for all $R(\nu) > 0$.

Since [20, p. 44]

$$(11) \qquad |J_\nu(z)| \leqslant \left|\frac{(z/2)^\nu}{\Gamma(\nu+1)}\right| \exp\frac{|z|^2}{4|\nu_0+1|},$$

where $|\nu_0 + 1|$ is the smallest of the numbers $|\nu+1|, |\nu+2|, \ldots,$ the series (10) is dominated, when $R(\nu) > 0$, by

$$\left|\frac{(\alpha z/2)^\nu}{\Gamma(\nu+1)}\right| \exp\frac{|z|^2}{4} \sum_{i=0}^{\infty} \frac{|1-\alpha^2|^i}{i!} \left|\frac{z}{2}\right|^{2i}$$

which obviously converges. The convergence is uniform with respect to z for z in any finite region, so our formal operation of integrating term by term with the aid of (6) is justified. It can be verified that the subsequent rearranging depends only on the absolute convergence of both (10) and the series obtained from (10) by differentiating. Requiring the coefficients of $J_\nu(z)$ and $J_{-\nu}(z)$ to vanish serves only to define a and b, and all the remaining coefficients vanish automatically. The convergence of (10) is also uniform with respect to ν, and hence we can let ν assume integral values. The formula (10) is therefore valid for all $R(\nu) \geqslant 0$.

For the solution of (5) to be $J_{-\nu}(\alpha z)$ ($R(\nu) > 0$ and ν not an integer), we find on equating coefficients of z^ν on both sides of (8) that $p_0 = 0$. Further, near $z = 0$,

$$J_{-\nu}(z) \approx \frac{1}{\Gamma(-\nu+1)}\left(\frac{z}{2}\right)^{-\nu} = \frac{\Gamma(\nu)\sin\nu\pi}{\pi}\left(\frac{z}{2}\right)^{-\nu}.$$

Using this result and (9) and putting $y = J_{-\nu}(\alpha z)$, we find, on equating coefficients of $(z/2)^{-\nu}$ in (8), that

$$\frac{\Gamma(\nu)\sin\nu\pi}{\pi}\alpha^{-\nu} = \frac{q_0}{\pi}\sum_{i=0}^{\infty}\frac{(-1)^i(1-\alpha^2)^i}{i!}\Gamma(\nu+i)\sin(\nu+i)\,\pi,$$

so that finally $q_0 = \alpha^\nu$ provided $|\, 1 - \alpha^2 \,| < 1$.

Using (11), we find that the series for $J_{-\nu}(\alpha z)$ is dominated by

$$\left|\frac{(2\alpha)^\nu\Gamma(\nu)\sin\nu\pi}{\pi z^\nu}\right| \exp\frac{|z|^2}{4|\nu_0+1|}\sum_{i=0}^{\infty}\frac{|1-\alpha^2|^i|\Gamma(\nu+i)|}{i!\,|\Gamma(\nu)|}$$

which converges if $|\, 1 - \alpha^2 \,| < 1$.

As before we therefore obtain, for $R(\nu) > 0$ and ν not an integer and $|\, 1 - \alpha^2 \,| < 1$,

$$J_{-\nu}(\alpha z) = \alpha^\nu\sum_{i=0}^{\infty}\frac{(1-\alpha^2)^i}{i!}\left(\frac{-z}{2}\right)^i J_{-\nu-i}(z).$$

Moreover,

$$\frac{J_\nu(\alpha z)\cos\nu\pi - J_{-\nu}(\alpha z)}{\sin\nu\pi}$$

$$= \alpha^\nu\sum_{i=0}^{\infty}\frac{(1-\alpha^2)^i}{i!}\left(\frac{z}{2}\right)^i\frac{J_{\nu+1}(z)\cos\nu\pi - (-1)^i J_{-\nu-i}(z)}{\sin\nu\pi}$$

$$= \alpha^\nu\sum_{i=0}^{\infty}\frac{(1-\alpha^2)^i}{i!}\left(\frac{z}{2}\right)^i\frac{J_{\nu+i}(z)\cos(\nu+i)\pi - J_{-\nu-i}(z)}{\sin(\nu+i)\,\pi};$$

that is,

$$Y_\nu(\alpha z) = \alpha^\nu \sum_{i=0}^{\infty} \frac{(1-\alpha^2)^i}{i!}\left(\frac{z}{2}\right)^i Y_{\nu+i}(z). \tag{12}$$

Equation (12) therefore holds, at least when $R(\nu) > 0$ and ν is not an integer, so long as $|1 - \alpha^2| < 1$ and z is finite and different from zero. That (12) holds under the same restrictions on α and z even when ν is an integer n, follows from the fact that

$$Y_{n+i}(z) = O((n+i)\,!)$$

(so that the series converges when $\nu = n$) and from the continuity of Y_ν as a function of ν.

Equations (10) and (12) are the so-called multiplication theorems for Bessel functions [20, p. 142].

EXPANSIONS OF THE CONFLUENT HYPERGEOMETRIC FUNCTIONS

We shall begin with the differential equation for the confluent hypergeometric functions $M_{k,\pm m}(y)$ given by Whittaker [21, p. 337]:

$$\frac{d^2W}{dy^2} + \left\{-\frac{1}{4} + \frac{k}{y} + \frac{\frac{1}{4} - m^2}{y^2}\right\} W = 0, \tag{13}$$

where $2m$ is assumed to be not an integer. The change of variables and parameters

$$4ky = z^2, \qquad W(y) = zF(z), \qquad 2m = \nu \tag{14}$$

transforms (13) into the form of (1):

$$\frac{d}{dz}\left(z\frac{dF}{dz}\right) + \left(z - \frac{\nu^2}{z}\right) F = \frac{z^3}{16k^2} F$$

where ν is not an integer, so that $v_1 = J_\nu(z)$, $v_2 = J_{-\nu}(z)$, and the corresponding integral equation for F becomes

$$F(z) = a\, J_\nu(z) + b\, J_{-\nu}(z) \tag{15}$$

$$- \frac{\pi}{32\, k^2 \sin \nu\pi} \int_\epsilon^z \{J_\nu(x)\, J_{-\nu}(z) - J_\nu(z)\, J_{-\nu}(x)\}\, x^3 F(x)\, dx.$$

Rearranging the terms in the series generated by the successive substitution procedure can again lead to a solution of the form (8). On putting (8) into the integral equation (15) we find, however, that before each integration can be performed with the aid of (6) and (7), it is necessary to use a recurrence relation on each term of the series.

Precisely, we rewrite (8) as

$$F(z) = \sum_{i=0}^{\infty} p_i z^i \left[2\frac{\nu + i + 1}{z} J_{\nu+i+1}(z) - J_{\nu+i+2}(z) \right]$$
$$+ \sum_{i=0}^{\infty} q_i z^i \left[-2\frac{\nu + i + 1}{z} J_{-\nu-i-1}(z) - J_{-\nu-i-2}(z) \right]$$

before substitution into the integral of (15). The relations (6) and (7) can then be applied. Proceeding formally, and putting $p_i = p_0 \bar{p}_i$, $q_i = q_0 \bar{q}_i$, $\bar{p}_0 = \bar{q}_0 = 1$ we obtain, for $i = 0, 1, 2, \ldots\ldots$, the relations

$$(16) \qquad \begin{cases} \bar{p}_1 = 0, \quad \bar{p}_2 = \dfrac{1}{32k^2}(\nu + 1), \\ \bar{p}_{i+3} = \dfrac{1}{32k^2}\left\{2\dfrac{\nu + i + 2}{i + 3}\bar{p}_{i+1} - \dfrac{\bar{p}_i}{i + 3}\right\} \end{cases}$$

and

$$(17) \qquad \begin{cases} \bar{q}_1 = 0, \quad \bar{q}_2 = \dfrac{1}{32k^2}(\nu + 1), \\ \bar{q}_{i+3} = \dfrac{1}{32k^2}\left\{2\dfrac{\nu + i + 2}{i + 3}\bar{q}_{i+1} + \dfrac{\bar{q}_i}{i + 3}\right\}. \end{cases}$$

Arguing as we did to establish the result (10) for $J_\nu(\alpha z)$, and returning to the original variables through (14), we obtain

$$(18) \qquad M_{k,m}(y) = \tfrac{1}{2} k^{-m-\frac{1}{2}} \Gamma(2m + 1) \sum_{i=0}^{\infty} \bar{p}_i (2\sqrt{ky})^{i+1} J_{2m+i}(2\sqrt{ky})$$

which is valid for all k, finite y, and for $R(m) \geqslant 0$; the $\bar{p}_i$ are given by (16). Using Langer's method, Taylor [18, Eq. (33)] has shown that the first term of this series is asymptotic to $M_{k,m}(y)$ as $|k| \to \infty$.

The corresponding result for $M_{k,-m}(y)$ with $R(m) > 0$ turns out to be an asymptotic expansion valid for large $|k|$. It is

$$(19) \qquad M_{k,-m}(y) \sim q_0 \sum_{i=0}^{\infty} \bar{q}_i (2\sqrt{ky})^{i+1} J_{-2m-i}(2\sqrt{ky})$$

for y finite and different from zero; the $\bar{q}_i$ are given by (17).

We shall first indicate how to prove that (19) is an asymptotic expansion valid for large $|k|$, and then find q_0 (see equation (21)). Substituting the expression

$$(20) \quad F(z) = p_0 \sum_{i=0}^{n} \bar{p}_i z^i J_{\nu+i}(z) + q_0 \sum_{i=0}^{n} \bar{q}_i z^i J_{-\nu-i}(z) + f(z)$$

into equation (15) and carrying out the integrations over the partial sums, we obtain an integral equation for $f(z)$. Using the recurrence relations (16) and (17) and choosing a, b or p_0, q_0 so that the coefficients of J_ν and $J_{-\nu}$ cancel out, we finally reduce the equation for $f(z)$ to the form

$$f(z) = \frac{p_0}{16k^2}\,[\bar{p}_{n-2}(\ldots.) + \bar{p}_{n-1}(\ldots.) + \bar{p}_n(\ldots.)]$$

$$+ \frac{q_0}{16k^2}\,[\bar{q}_{n-2}(\ldots)\ldots] - \frac{\pi}{32k^2\sin\nu\pi}\int_\epsilon^z \{J_\nu(x)J_{-\nu}(z) - J_\nu(z)J_{-\nu}(x)\}$$

$$x^3 f(x)\,dx.$$

Now, if M denotes the maximum of $|f(x)|$ along the path of integration, then

$$M \leqslant (|p_0|A + |q_0|B + MC)/|k|^2$$

where A, B, C are upper bounds for corresponding expressions in the preceding equation. Therefore

$$M \leqslant \frac{|p_0|A + |q_0|B}{|k|^2(1 - C/|k|^2)}$$

if $|k|$ is chosen large enough that the denominator is positive. From (16) and (17), both $\bar{p}_\iota$, $\bar{q}_\iota$ are $O(|k|^{-2\iota/3})$, so that

$$M = p_0\,O(|k|^{-2(n+1)/3}) + q_0\,O(|k|^{-2(n+1)/3});$$

moreover, the last term in each of the partial sums on the right of (20) is $O(|k|^{-2n/3})$. The series obtained from (20) is therefore asymptotic to the solution of (15) for large $|k|$.

Equating coefficients of z^ν and $z^{-\nu}$ in the $F(z)$ corresponding to $M_{k,-m}(y)$ [21, p. 338] and in the right side of (20), we obtain

$$p_0 = 0$$

$$(4k)^{(\nu-1)/2} \sim q_0(\sin\nu\pi/\pi)\sum_{\iota=0}^{\infty}(-1)^\iota\bar{q}_\iota\Gamma(\nu+\iota).$$

Hence, returning to the original variables, we obtain an asymptotic expansion for the coefficient q_0 in (19):

$$(21)\qquad 1/q_0 \sim (4k)^{-m+\frac{1}{2}}(\sin 2m\pi/\pi)\sum_{\iota=0}^{\infty}(-1)^\iota\bar{q}_\iota\Gamma(2m+\iota).$$

Appropriate changes of variables in (18) and (21) lead to expansions of, for example, the Laguerre and the Hermite functions.

CONCLUDING REMARKS

We have shown that the usual method of studying the asymptotic solutions of differential equations can, with some modification, be used to obtain complete expansions of special functions in certain important cases. In a future paper we propose to adapt the method to the solution of certain boundary value problems—the so-called bounded quantum mechanical problems.

Kuhn [13] presents considerable analytic and numerical detail obtained from expansions of the confluent hypergeometric functions in series of Bessel functions. His expansions can be obtained from the integral equation (15) by successive substitutions without any subsequent rearrangement of terms except for collecting the J_ν and $J_{-\nu}$ terms.

Using a variety of methods, other expansions of the confluent hypergeometric functions have been found by Abramowitz [1], Buchholz [4], Henrici [9], Karlin [12], Pignedoli [17], and Tricomi [19].

Asymptotic expansions of the confluent hypergeometric functions are given or referred to by Chang, Chu and O'Brien [6]. Such expansions have also been considered by Moecklin [16] and Buchholz [5].

We wish to express our appreciation of many interesting and helpful discussions with Dr. G. E. Latta.

REFERENCES

1. M. Abramowitz, J. Math. Phys., vol. 29 (1950), 303–8.
2. G. D. Birkhoff, Trans. Amer. Math. Soc., vol. 9 (1908), 219–31.
3. ——— Trans. Amer. Math. Soc., vol. 9 (1908), 373–95.
4. H. Buchholz, Math. Z., vol. 53 (1950), 387–402.
5. ——— Z. angew. Math. Mech., vol. 30 (1950), 133–48.
6. C. Chang, B. Chu, and V. O'Brien, J. Rat. Mech. Anal., vol. 2 (1953), 125–35.
7. T. M. Cherry, Trans. Amer. Math. Soc., vol. 68 (1950), 224–57.
8. A. S. Eddington, *The mathematical theory of relativity*, 2nd ed. (Cambridge, 1924), p. 90.
9. P. Henrici, Proc. Amer. Math. Soc., vol. 4 (1953), 329–34.
10. E. L. Ince, *Ordinary differential equations*, 1st Amer. ed. (Dover, 1926), p. 271.
11. Y. Ikeda, Math. Z., vol. 22 (1925), 16–25.
12. M. Karlin, J. Math. Phys., vol. 28 (1949), 13–4.
13. T. S. Kuhn, Quart. Appl. Math., vol. 9 (1951), 1–16.
14. R. E. Langer, Trans. Amer. Math. Soc., vol. 67 (1949), 461–90.
15. K. S. Miller, J. Appl. Phys., vol. 22 (1951), 1054–7.
16. E. Moecklin, Comment. Math. Helv., vol. 7 (1934–5), 24–6.
17. A. Pignedoli, Atti. Sem. Mat. Fis. Univ. Modena, vol. 4 (1950), 3–25.
18. W. C. Taylor, J. Math. Phys., vol. 18 (1939), 34–49.
19. F. Tricomi, Giorn. Ist. Ital. Attuari, vol. 12 (1942), 14–33.
20. G. N. Watson, *Theory of Bessel functions*, 2nd ed. (Cambridge, 1952).
21. E. T. Whittaker and G. N. Watson, *Modern analysis*, 4th ed. (Cambridge, 1952).

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLVII : SERIES III : JUNE, 1953
SECTION THREE

The Ergodic Theorem for Banach Spaces with Convex-Compactness

VALENTINA KROTKOV and ISRAEL HALPERIN, F.R.S.C.

1. INTRODUCTION

LET T be a bounded, linear operator on a real or complex Banach space B and let

$$T_n = \frac{1}{n+1}(1 + T + T^2 + \ldots + T^n).$$

T is said to be ergodic on B in the sense of J. von Neumann if for each f in B, $T_n f$ converges to a limit when n becomes infinite.

In order that T be ergodic on B, the following conditions are necessary:

(1.1) $(1/n)\ T^n f \rightarrow 0$ when $n \rightarrow \infty$, for every f in B,

(1.2) $| T_n | \leqslant K$ for all n, for some $K < \infty$.

Indeed, (1.1) follows from the relation

$$(1.3) \qquad (1/n)\ T^n = \frac{n+1}{n} T_n - T_{n-1}$$

and (1.2) from the theorem that a convergent sequence of bounded, linear operators must be uniformly bounded [1, p. 80, théorème 5].

Clearly, if T is ergodic on B and $T_n f$ converges to f^*, say, then (1.1) and the identity

$$T_n(Tf) = T(T_n f) = T_n f + \frac{1}{n+1}(T^{n+1}f - f)$$

together with the continuity of T, show that $Tf^* = f^*$.

Because of (1.3), (1.2) implies that $| (1/n)T^n | \leqslant 3K$ for all n and hence that $(1/n)T^n f$ will converge to 0 for all f if $(1/n)T^n f$ converges to zero for a set of f which span B, that is, whose linear combinations are dense in B.

Both (1.1) and (1.2) are implied by the condition:

$$(1.4) \qquad | T^n | \leqslant K \text{ for all } n, \text{ for some } K < \infty.$$

But if T is ergodic on B and (1.4) holds, then the identity

$$T_n(T^r f) - f^* = T^r(T_n f - f^*)$$

shows that $T_n(T^r f)$ converges to f^*, irrespective of how the non-negative integer r varies, when n becomes infinite.

In this note we discuss a method of F. Riesz [6; 7] to show that the conditions (1.1) and (1.2) are sufficient to imply that T is ergodic on B if B is *convex-compact.* By definition, we shall say that a subset A of B (possibly all of B) is convex-compact if every decreasing sequence of non-empty, closed, bounded, *convex* subsets of A has a non-empty intersection.

2. THE RIESZ METHOD OF AVERAGES

Suppose that B is an arbitrary Banach space and that T satisfies (1.1) and (1.2). Let $S(f)$ denote the set of all "averages" of the form

$$g = c_0 f + c_1 Tf + \ldots + c_r T^r f$$

where r is an arbitrary non-negative integer and the c_i are non-negative scalars with sum equal to 1. Riesz observed that for such g and for $n > r$,

$$T_n f - T_n g = \frac{1}{n+1} \sum_{j=0}^{r-1} (c_{j+1} + \ldots + c_r)(T^j f - T^{n+1+j} f).$$

Hence (1.1) implies

$$\overline{\lim_{n\to\infty}}\, |T_n f - T_n g| = 0,$$

and therefore

$$\begin{aligned}\overline{\lim_{n,m\to\infty}}\, |T_n f - T_m f| &\leqslant \overline{\lim_{n\to\infty}}\, |T_n f - T_n g| + \overline{\lim_{n\to\infty}}\, |T_n g - g| \\ &\quad + \overline{\lim_{m\to\infty}}\, |g - T_m g| + \overline{\lim_{m\to\infty}}\, |T_m g - T_m f| \\ &= 2\, \overline{\lim_{n\to\infty}}\, |T_n g - g|.\end{aligned}$$

Thus, following Riesz, T will be ergodic on B if there are elements g_i in $S(f)$ such that

(2.1) $$\overline{\lim_{n\to\infty}}\, |T_n g_i - g_i| \to 0 \qquad \text{when } i \to \infty.$$

Now there will be such g_i in $S(f)$ if there is an element h in $\overline{S(f)}$, the closure of $S(f)$, with $Th = h$. Indeed, there will then be g_i in $S(f)$ such that $|\, g_i - h \,| \to 0$ when $i \to \infty$; since

$$T_n g_i - g_i = T_n(g_i - h) + (T_n h - h) + (h - g_i),$$

the identity $T_n h = h$ together with (1.2) shows that (2.1) will hold.

Thus T will be ergodic on B if for every f, the mapping of $\overline{S(f)}$ into itself by T has a fix-point.

Now let $S_n(f)$ be the set of elements u in $\overline{S(f)}$ satisfying both $| u | \leqslant K | f |$ and $| Tu - u | = | (T - 1)u | \leqslant 1/n$. Then every $S_n(f)$ is non-empty; for $| T_m f | \leqslant K | f |$ for all m by (1.2), and the identity

$$\left| T(T_m f) - T_m f \right| = \frac{1}{m+1} \left| T^{m+1}f - f \right| \leqslant \frac{1}{m+1} \left| T^{m+1}f \right| + \frac{1}{m+1} \left| f \right|$$

together with (1.1) show that, for fixed n, $T_m f$ is contained in $S_n(f)$ for all sufficiently large m. It is easily verified that for fixed f, the $S_n(f)$ are a non-increasing sequence of bounded, closed, convex sets. If B is convex-compact then some element h will be in all $S_n(f)$; clearly, this h will be in $\overline{S(f)}$ and for every n, $| Th - h | \leqslant 1/n$ so that $Th = h$.

This proves: *if B is convex-compact then a bounded, linear operator T is ergodic on B if and only if T satisfies* (1.1) *and* (1.2).

Since metrically closed convex subsets of B are necessarily weakly closed (this is a theorem of Ascoli-Mazur, see Halperin, these Transactions, vol. 47, pp. 1–6) weak convex-compactness is equivalent to the strong convex-compactness as defined above. Hence our ergodic theorem includes the case of locally weakly compact Banach spaces [8] and the equivalent case of reflexive Banach spaces [5], equivalent by the theorem of Eberlein. These writers assume (1.4) but their arguments are actually valid assuming only (1.1) and (1.2). The reflexive spaces are now known to include the uniformly convex Banach spaces of Clarkson [2], and in particular, the L^p spaces, $p > 1$. We refer to [8] for references to other writers on this subject but particular reference should be made to the proof of an ergodic theorem by Riesz [7].

3. THE UNIFORMLY CONVEX BANACH SPACE

We note that it is particularly easy to show that every uniformly convex B is convex-compact. Indeed, if S_n are a decreasing sequence of non-empty, bounded, closed, convex sets, let μ_n = infimum $| g |$ for all g in S_n, and let g_n be selected from S_n with $\mu_n \leqslant | g_n | \leqslant \mu_n + 1/n$. Then the μ_n are non-decreasing and bounded above since the S_n are bounded. It follows that $| g_n |$ converges to a limit as n becomes infinite and the convexity of the S_n together with the uniform convexity of B implies that the g_n converge to a limit h. This h is necessarily in all S_n, assumed closed.

4. ADDITIONAL COMMENTS

Call an element f in the Banach space T-ergodic if $T_n f$ converges when n becomes infinite. Then (1.2) implies that the T-ergodic elements form a closed linear subspace of B. Our argument actually shows that f is T-ergodic if $\overline{S(f)}$ is convex-compact, and that T is ergodic on B if $\overline{S(f)}$ is convex-compact for a set of f which span B, assuming (1.1) and (1.2).

Finally, consider the special case that B is L^1: the space of $f(P)$ summable on a set X of points P with a non-negative, countably additive measure $|e|$, and that T is a bounded, linear operator of the type

$$Tf(P) = f(\phi(P))$$

for some mapping $\phi(P)$ of X into itself. It is easy to see that L^1 is not convex-compact, except when X consists essentially of a finite number of points. However, if X has finite measure and T is of this special type then for every measurable subset e the function 1_e, which is defined to have value 1 if P is in e, and value 0 otherwise, has $S(1_e)$ convex-compact, in fact every sequence of elements g_n from $S(1_e)$ has a weakly convergent subsequence. This follows from the fact that the g_n are uniformly bounded ($|g_n(P)| \leqslant 1$ for all P) and X is of finite measure. Thus (1.1) and (1.2) imply that such a T is ergodic on such an L^1(see [4]); but (1.1) may be omitted since it is automatically valid for all 1_e when X is of finite measure, and the 1_e span L^1; also, (1.2) can be deduced if $|T_n 1_e| \leqslant K|e|$ is known to be valid for all e, as pointed out by Dunford and Miller [3]. Thus the necessary and sufficient condition that a T of this type be ergodic on such an L^1 is that for some $K < \infty$ and for all n and all e, (see [3]),

$$\frac{1}{n+1}(|e| + |\phi^{-1}(e)| + \ldots + |\phi^{-n}(e)|) \leqslant K|e|.$$

REFERENCES

1. S. Banach, *Théorie des opérations linéaires* (Warsaw, 1932).
2. J. A. Clarkson, *Uniformly convex spaces*, Trans. Amer. Math. Soc., vol. 10 (1936), 396–414.
3. N. Dunford and D. S. Miller, *On the ergodic theorem*, Trans. Amer. Math. Soc., vol. 60 (1946), 538–49.
4. S. Kakutani, *Mean ergodic theorems in abstract (L)-spaces*, Proc. Imp. Acad. Tokyo, vol. 15 (1939), 121–3.
5. E. R. Lorch, *Means of iterated transformations in reflexive vector spaces*, Bull. Amer. Math. Soc., vol. 45 (1939), 945–7.
6. F. Riesz, *Sur la théorie ergodique*, Comment. Math. Helv., vol. 17 (1945), 221–39.
7. ——— *Some mean ergodic theorems*, J. London Math. Soc., vol. 13 (1938), 274–8.
8. K Yosida and S. Kakutani, *Operator-theoretical treatment of Markoff's process and mean ergodic theorem*, Ann. Math., vol. 42 (1941), 188–228.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLVII : SERIES III : JUNE, 1953
SECTION THREE

A Problem in Combinatorial Analysis

N. S. MENDELSOHN

Presented by R. L. JEFFERY, F.R.S.C.

1. INTRODUCTION

PROBLEMS in combinatorial analysis are usually considered solved when one can find either a difference equation or a recurrence relationship which enables one to compute $f(n)$ step by step. If it makes computation easier, an explicit formula for $f(n)$ is a more desirable type of solution. In common with other branches of number theory, certain published "solutions" of combinatorial problems merely replace the original problem by another one which may in fact be more difficult to solve than the original problem. The author of this paper has discovered in [1] and [2] a symbolic method of setting up and solving the recurrence relations for a wide class of combinatorial problems. The problem studied in this paper can be handled by the author's general method, but the computations are very laborious. However, the formula, once obtained, can be verified much more easily and in this paper we will present only this easy verification. Numerical examples will illustrate the computational efficiency of our formula.

2. THE PROBLEM

We will consider the following problem. Suppose we have a set of $a + b + c + \dots$ objects of which a are alike of one kind, b alike of a second kind, etc. Let $C_r[a, b, c, \dots]$ denote the number ofcombinations taken r at a time of these objects. We will write $C_r[a^3, b^2, c, \dots]$ as an abbreviation for $C_r[a, a, a, b, b, c, \dots]$. An explicit formula for

$$C_r[a_1^{k_1}, a_2^{k_2}, \dots, a_t^{k_t}]$$

will be obtained. For the corresponding permutation problem, i.e., the problem of finding the number of permutations r at a time of the given class of objects, a recurrence formula will be obtained. Corresponding to the notation $C_r[a, b, c, \dots]$ we will use $P_r[a, b, c, \dots]$ in the case of the permutation problem.

3. The Fundamental Recurrence

We first show that

$$C_r[a_1^{k_1}, a_2^{k_2}, \ldots, a_t^{k_t}]$$

satisfies the following recurrence relationship:

$$\text{(I)} \qquad C_r[a_1^{k_1}, a_2^{k_2}, \ldots, a_t^{k_t}] = \sum_{u=0}^{\min(r, a_t)} C_{r-u}[a_1^{k_1}, a_2^{k_2}, \ldots, a_t^{k_t-1}]$$

To prove this formula we isolate one set S of a_t identical objects and observe how these objects enter into the combinations. The combinations which are counted by the expression

$$C_r[a_1^{k_1}, a_2^{k_2}, \ldots, a_t^{k_t}]$$

can be divided into the following mutually exclusive classes:

(0) Those which contain no object of the set S. There are

$$C_r[a_1^{k_1}, a_2^{k_2}, \ldots, a_t^{k_t-1}]$$

of these.

(1) Those which contain one object of the set S. There are

$$C_{r-1}[a_1^{k_1}, a_2^{k_2}, \ldots, a_t^{k_t-1}]$$

of these.

. . . .

(u) Those which contain u objects of the set S. There are

$$C_{r-u}[a_1^{k_1}, a_2^{k_2}, \ldots, a_t^{k_t-1}]$$

of these.

. . . .

On adding the expressions obtained in (0), (1), . . . , (u), . . . we obtain formula (I).

Formula (I) is quite efficient for computing purposes but an explicit expression will be given in the next section which is actually much better.

4. The Explicit Formula

An explicit formula for

$$C_r[a_1^{k_1}, a_2^{k_2}, \ldots, a_t^{k_t}]$$

can now be obtained. However, we will write down the formula for the case $C_r[a^i, b^j]$ only. A simple examination of this formula will show how it may be extended to the general case. The formula in question is:

$$\text{(II)}\quad C_r[a^i, b^j] = \sum_k \sum_m (-1)^{k+m}\binom{i}{k}\binom{j}{m} \binom{i+j+r-1-(a+1)\,k-(b+1)\,m}{r-(a+1)\,k-(b+1)\,m}$$

where the summation is taken over all integral values of the indices k, m satisfying the inequalities: $0 \leqslant k \leqslant i, 0 \leqslant m \leqslant j, 0 \leqslant (a+1)k + (b+1)m \leqslant r$. In practice, it will be seen that these inequalities confine k and m to a very small range. In the next section we will show that formula (II) is quite easy to handle numerically. To prove formula (II), we first note that if $j = 0$, (II) is replaced by the formula

$$\text{(III)}\qquad C_r[a^i] = \sum_k (-1)^k \binom{i}{k}\binom{i+r-1-(a+1)\,k}{r-(a+1)\,k}$$

summed over all values of k in the interval

$$0 \leqslant k \leqslant \min\left(i, \left[\frac{r}{a+1}\right]\right)$$

where $[X]$ stands for the greatest integer less than or equal to X. When $i = 1$, (III) becomes $C_r[a] = 1$, if $r \leqslant a$, and $C_r[a] = 0$ for $r > a$. This result is obviously true, so that (III) is true if $i = 1$. The recurrence (I) for the case $t = 1$, $k_1 = i$ is

$$\text{(IV)}\qquad C_r[a^i] = \sum_{u=0}^{\min(r,a)} C_{r-u}[a^{i-1}].$$

Assuming (III) to be true when i is replaced by $i - 1$, and using (IV) one obtains that (III) is true for i if it is true for $i - 1$. Hence (III) is true for all integral i, by induction. Now $C_r[a^i] = C_r[a_i, b^0]$ so that (II) is true when $j = 0$. By a similar application of (I) to the case $t = 2$ we can establish (II) by induction, to be true for all j. This completes the proof.

One can obtain simpler formulae in case any of the a_i are greater than or equal to r. In this case the number of combinations is actually independent of the exact value of a_i. We will replace a_i by the symbol ∞ in case $a_i \geqslant r$ so that, for example, $C_5[2^3, 3^2, 5^4, 6^8, 7^9]$ will be written as $C_5[2^3, 3^2, \infty^{21}]$. The formula for $C_r[a^i, b^j, \infty^p]$ can be verified to be

$$\text{(V)}\quad C_r[a^i, b^j, \infty^p] = \sum_k \sum_m (-1)^{k+m}\binom{i}{k}\binom{j}{m} \binom{i+j+r+p-1-(a+1)\,k-(b+1)\,m}{r-(a+1)\,k-(b+1)\,m}$$

where the summation indices k, m range over all integers satisfying $0 \leqslant k \leqslant i$, $0 \leqslant m \leqslant j$, $0 \leqslant (a+1)k + (b+1)m \leqslant r$.

In the next section we will illustrate the efficacy of these formulae by numerical examples.

5. Some Illustrative Examples

As a first example we will consider the number of essentially different 11-card canasta hands which can be formed from a 104-card pack. (We do not count the red threes as part of the pack since these are always replaced in a hand.) The pack consists of four black threes, four jokers, eight aces, eight twos, eight fours, ... , eight kings. Cards of a given denomination are considered equivalent regardless of suit so that the number of ways of choosing eleven cards from such a pack is $C_{11}[8^{12}, 4^2]$. Substituting in (II) we obtain

$$C_{11}[8^{12}, 4^2] = \sum_k \sum_m (-1)^{k+m} \binom{12}{k} \binom{2}{m} \binom{24 - 9k - 5m}{11 - 9k - 5m}$$

summed over all pairs of integers (k, m) subject to the inequalities $0 \leqslant k \leqslant 12$, $0 \leqslant m \leqslant 2$, $0 \leqslant 9k + 5m \leqslant 11$. The pairs (k, m) satisfying these inequalities are: $(0, 0)$, $(0, 1)$, $(0, 2)$, $(1, 0)$. Hence,

$$C_{11}[8^{12}, 4^2] = \binom{24}{11} - 2\binom{19}{6} + 14 - 12\binom{15}{2} = 2{,}440{,}634.$$

It will be noted that our solution requires the computation of four terms only. A solution based on the partitions of eleven would require the computation of 52 terms each of which is a product of binomial coefficients. A solution by J. P. Ballantyne [3] would require the construction of a table containing 12 rows and 16 columns, i.e., one containing 192 entries.

As a second example, we consider the problem of finding the number of essentially different bridge hands. By this we mean the number of different bridge hands which can be formed when the cards in each suit from 2 to 9 inclusive are ranked equally. The required number is $C_{13}[8^4, 1^{20}]$. By (II)

$$C_{13}[8^4, 1^{20}] = \sum_k \sum_m \binom{4}{k} \binom{20}{m} \binom{36 - 9k - 2m}{13 - 9k - 2m} (-1)^{k+m}$$

summed over all pairs (k, m) such that $0 \leqslant k \leqslant 4$, $0 \leqslant m \leqslant 20$, $0 \leqslant 9k + 2m \leqslant 13$. Hence, the admissible (k, m) are: $(0, 0)$, $(0, 1)$, $(0, 2)$, $(0, 3)$, $(0, 4)$, $(0, 5)$, $(0, 6)$, $(1, 0)$, $(1, 1)$, $(1, 2)$. This yields

$$\begin{aligned} C_{13}[8^1, 1^{20}] &= \binom{36}{13} - 20\binom{34}{11} + 190\binom{32}{9} - 1140\binom{30}{7} + 4845\binom{28}{5} \\ &\quad - 15504\binom{26}{3} + 38760\binom{24}{1} - 4\binom{27}{4} + 80\binom{25}{2} - 760 \\ &= 34{,}033{,}880. \end{aligned}$$

Here again, the saving in computation is very considerable when compared with the usual counting methods. This result is in agreement with that given by Ballantyne in [3] and affords a comparison between the two methods.

6. The Corresponding Permutation Problem

We will denote by $P_r[a, b, c, \dots]$ the number of permutations taken r at a time of $a + b + c + \dots$ objects of which a are alike of one kind, b are alike of another kind, etc. We will use the notation

$$P_r[a_1^{k_1}, a_2^{k_2}, \dots, a_t^{k_t}]$$

$$\text{and } P_r[a^i, \infty^j]$$

to represent the permutation analogues of

$$C_r[a_1^{k_1}, a_2^{k_2}, \dots, a_t^{k_t}] \text{ and } C_r[a^i, \infty^j].$$

We will confine ourselves mainly to the problem of finding the permutation analogue of (I) namely:

$$\text{(VI)} \quad P_r[a_1^{k_1}, a_2^{k_2}, \dots, a_t^{k_t}] = \sum_{u=0}^{\min(r,a_t)} \binom{r}{u} P_{r-u}[a_1^{k_1}, a_2^{k_2}, \dots, a_t^{k_t-1}]$$

Before proving (VI) we note that the formulae $P_r[\infty^s] = s^r$ and

$$P_r[1^s] = \frac{s!}{(s-r)!}$$

are easily obtained from (VI) by induction. These formulae are, in fact, classical results.

The proof of (VI) is analogous to the proof of (I). We isolate one set S of a_t identical objects and then divide the totality of permutations into the following exclusive classes: For $\nu = 0, 1, 2, 3, \dots,$ $\min(r, a_t)$, we form the class $P_{(\nu)}$ of all permutations containing exactly ν elements of the set S. Then, obviously,

$$P_r[a_1^{k_1}, a_2^{k_2}, \dots, a_t^{k_t}] = \sum_{\nu=0}^{\min(r,a_t)} P_{(\nu)}.$$

To compute $P_{(\nu)}$ we note that each permutation in the set $P_{(\nu)}$ can be obtained by forming a permutation from $r - \nu$ objects chosen from the totality of objects except those in S and then introducing into this permutation ν objects taken from S. The number of ways of introducing these ν objects into the permutation is

$$\binom{r}{\nu},$$

and the number of permutations into which these objects must be introduced is

$$P_{r-\nu}[a_1^{k_1}, a_2^{k_2}, \dots, a_t^{k_t-1}].$$

Hence,

$$P_{(\nu)} = \binom{r}{\nu} P_{r-\nu}[a_1^{k_1}, a_2^{k_2}, \dots, a_t^{k_t-1}].$$

This completes the proof of formula (VI).

As yet, the author has been unable to transform the recurrence (VI) into an explicit formula. In spite of this, formula (VI) gives rise to a rapid computational procedure, and in special cases does lead to explicit formulae.

REFERENCES

1. N. S. MENDELSOHN, *Symbolic solution of card matching problems*, Bull. Amer. Math. Soc., vol. 52 (1946), 718–924.
2. ——— *Applications of combinational formulae to generalizations of Wilson's Theorem*, Can. J. Math., vol. 1 (1949), 328–36.
3. J. P. BALLANTYNE: *$C_{n,r}$ where the n things include identities*, Amer. Math. Monthly, vol. 57 (1950), 626–9.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLVII : SERIES III : JUNE, 1953
SECTION THREE

Note on a Combinatorial Formula of Mendelsohn

LEO MOSER

Presented by R. L. JEFFERY, F.R.S.C.

THE object of this note is to give a short, direct proof of result (II) of the preceding paper. The result in question is

$$\text{(II)} \quad C_r\,[a^i, b^j] = \sum_k \sum_m (-1)^{k+m}\binom{i}{k}\binom{j}{m} \binom{i+j+r-1-(a+1)\,k-(b+1)\,m}{r-(a+1)\,k-(b+1)\,m}$$

where the summation ranges over all values of k and m for which the binomial coefficients involved are positive.

Using the set up of Ballantyne's paper [3], we have $C_r[a^i, b^j]$ is the coefficient of x^r in

$$\begin{aligned} f(x) &= (1+x+x^2+\ldots+x^a)^i(1+x+x^2\ldots+x^b)^j \\ &= (1-x^{a+1})^i\,(1-x^{b+1})^j\,(1-x)^{-(i+j)} \end{aligned}$$

Now

$$(1-x)^{-(i+j)} = \sum_{t>0}\binom{i+j-1+t}{t}x^t$$

Hence

$$C_r\,[a^i, b^j] = \sum_k \sum_m (-1)^{k+m}\binom{i}{k}\binom{j}{m}\binom{i+j-1+t}{t}$$

where t is determined by $r=(a+1)k+(b+1)m+t$. This completes the proof of (II).

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLVII : SERIES III : JUNE, 1953
SECTION THREE

Some Remarks on Laplace's Method

P. G. ROONEY

Presented by M. WYMAN, F.R.S.C.

IN his book, *The Laplace Transform*, D. V. Widder finds asymptotic formulae for integrals of the form

$$\int_a^b \exp(kh(x))\,\phi(x)\,dx$$

as $k \to \infty$, under the conditions that $h(x)$ has a simple flat maximum at a point $x = c$ between a and b, i.e. $h'(c) = 0$, $h''(c) < 0$, and $\phi(x)$ has, at worst, a simple zero at $x = c$ and does not change sign there, together with certain other conditions.

These results can easily be generalized to the case in which $h(x)$ has a higher order maximum at $x = c$, and $\phi(x)$ has a zero of arbitrary order there, though $\phi(x)$ is still not allowed to change sign at $x = c$. The generalization is contained in the following two theorems.

The first theorem deals with the case in which the maximum of $h(x)$ occurs at one end point of the interval, and the second theorem with the case in which the maximum of $h(x)$ occurs at an interior point of the interval.

THEOREM 1. *If*

(1) $a < a + \eta < b$,

(2) $h(x) \in C^{2m}(a \leqslant x \leqslant a + \eta)$, $h^{(n)}(a) = 0$, $(n = 1, 2, \ldots, 2m-1)$, $h^{(2m)}(a) < 0$, $h(x)$ *is non-increasing in* $a \leqslant x \leqslant b$,

(3) $(x - a)^\nu \phi(x) \in L(a, b)$, $\phi(a+)$ *exists and* $\phi(a+) \neq 0$,

then

$$\int_a^b \exp(kh(x))(x-a)^\nu \phi(x)\,dx$$
$$\sim (-(2m)!/kh^{(2m)}(a))^{(\nu+1)/2m}(\phi(a+)/2m)\exp(kh(a))\Gamma((\nu+1)/2m)$$

as $k \to \infty$.

Proof. We may assume $\phi(a+) > 0$. Then for every ϵ such that $0 < \epsilon < \phi(a+)$ and $0 < \epsilon < -h^{(2m)}(a)$, δ exists such that

(i) $0 < A_1 = \phi(a+) - \epsilon < \phi(x) < A_2 = \phi(a+) + \epsilon$,

(ii) $B_1 = h^{(2m)}(a) - \epsilon < h(x) < B_2 = h^{(2m)}(a) + \epsilon < 0$, for all x, $a \leqslant x \leqslant a + \delta$.

Let

$$I_k = \int_a^b \exp\{k(h(x) - h(a))\}(x-a)^\nu \phi(x)\,dx$$

$$= \int_a^{a+\delta} + \int_{a+\delta}^b \equiv I_k' + I_k''.$$

Then

$$|I_k''| \leqslant \exp\{k(h(a+\delta) - h(a))\}\int_a^b |\phi(x)|(x-a)^\nu\,dx.$$

Since $h(a+\delta) - h(a) < 0$, it is clear that $I_k'' = O(\alpha^k)$, $0 < \alpha < 1$.

Using (i) and Taylor's theorem with remainder, we have

$$A_1\int_a^{a+\delta} \exp\{kh^{(2m)}(\xi)(x-a)^{2m}/(2m)!\}(x-a)^\nu\,dx < I_k'$$

$$< A_2\int_a^{a+\delta} \exp\{kh^{(2m)}(\xi)(x-a)^{2m}/(2m)!\}(x-a)^\nu\,dx,$$

where $a < \xi < a + \delta$. From (ii) we only strengthen these inequalities by replacing $h^{(2m)}(\xi)$ by $h^{(2m)}(a) - \epsilon$ on the right and by $h^{(2m)}(a) + \epsilon$ on the left. Thus

$$A_1\int_a^{a+\delta} \exp\{kB_1(x-a)^{2m}/(2m)!\}(x-a)^\nu\,dx < I_k'$$

$$< A_2\int_a^{a+\delta} \exp\{kB_2(x-a)^{2m}/(2m)!\}(x-a)^\nu\,dx.$$

In the left hand side of these inequalities let

$$-u = kB_1(x-a)^{2m}/(2m)!,$$

and in the right hand side, let

$$-u = kB_2(x-a)^{2m}/(2m)!.$$

Thus the inequalities become

$$(-(2m)!/kB_1)^{(\nu+1)/2m}(A_1/2m)\int_0^{kC_1} e^{-u}u^{((\nu+1)/2m)-1}\,du < I_k'$$

$$< (-(2m)!/kB_2)^{(\nu+1)/2m}(A_2/2m)\int_0^{kC_2} e^{-u}u^{((\nu+1)/2m)-1}\,du$$

where $C_i = -B_i\delta^{2m}/(2m)!$ $(i = 1, 2)$.

Thus,

$$(-(2m)!/B_1)^{(\nu+1)/2m}(A_1/2m)\,\Gamma((\nu+1)/2m) \leqslant \overline{\underline{\lim}}\, I_k'$$

$$\leqslant (-(2m)!/B_2)^{(\nu+1)/2m}(A_2/2m)\,\Gamma((\nu+1)/2m).$$

Hence, since ϵ is arbitrary,

$$\lim_{k\to\infty} k^{(\nu+1)/2m} I_k = \lim_{k\to\infty} k^{(\nu+1)/2m} I_k'$$
$$= (-(2m)!/h^{(2m)}(a))^{(\nu+1)/2m}(\phi(a+)/2m)\ \Gamma((\nu+1),2m).$$

COROLLARY 1. *If*

(1) $a < b - \eta < b$,

(2) $h(x) \in C^{2m}(b-\eta \leqslant x \leqslant b), h^{(n)}(b) = 0, (n = 1, 2, \ldots, 2m-1)$, $h^{(2m)}(b) < 0$, *$h(x)$ is non-decreasing in $a \leqslant x \leqslant b$,*

(3) $(b-x)^\nu \phi(x) \in L(a,b)$, *$\phi(b-)$ exists and $\phi(b-) \neq 0$,*

then,

$$\int_a^b \exp\ (kh(x))(b-x)^\nu\ \phi(x)\,dx$$
$$\sim(-(2m)!/kh^{(2m)}(b))^{(\nu+1)/(2m)}(\phi(b-)/2m)\exp\ (kh(b))\ \Gamma((\nu+1)/2m)$$

as $k \to \infty$.

COROLLARY 2. *If* (1), (2), *and* (3) *of Theorem 1 hold, with the exception that $\phi(a+) = 0$, then*

$$\int_a^b \exp\ (kh(x))(x-a)^\nu\ \phi(x)\,dx = o(k^{-(\nu+1)/2m}\exp\ (kh(a))).$$

THEOREM 2. *If*

(1) $a < b - \eta < b < b + \eta < c$,

(2) $h(x) \in C^{2m}(b-\eta \leqslant x \leqslant b+\eta)$, $h^{(n)}(b) = 0$, $(n = 1, 2, \ldots, 2m-1)$, $h^{(2m)}(b) < 0$, *$h(x)$ is non-decreasing in $a \leqslant x \leqslant b$, and non-increasing in $b \leqslant x \leqslant c$,*

(3) $|x-b|^\nu\ \phi(x) \in L(a,c)$, *$\phi(b+)$, $\phi(b-)$ exist and are not zero.*

then

$$\int_a^c \exp\ (kh(x))|x-b|^\nu\ \phi(x)\,dx$$
$$\sim(-(2m)!/kh^{(2m)}(b))^{(\nu+1)/2m}((\phi(b+)+\phi(b-))/2m)$$
$$\exp\ (kh(b))\ \Gamma((\nu+1)/2m)$$

as $k \to \infty$.

Proof. If we write the integral as the integral from a to b plus the integral from b to c, and apply Corollary 1 to the first and Theorem 1 to the second integral, we arrive at the result directly.

As an example, we may find the asymptotic value of

$$\int_0^1 e^{-kx^2}\sin^6 x\,dx$$

This integral fulfills all the hypothesis of Theorem 1, since $\sin^6 x$ may be considered as $x^6\,\phi(x)$ with $\phi(0+) = 1$. Thus we have

$$\int_0^1 e^{-kx^8}\sin^6 x\,dx \sim \tfrac{1}{8}k^{-7/8}\,\Gamma(\tfrac{7}{8}) \qquad \text{as } k \to \infty.$$

Further, since

$$\int_1^\infty e^{-kx^8}\sin^6 x\,dx\Bigg| \leqslant e^{-(k-1)}\int_1^\infty e^{-x^8}\,dx,\ k \geqslant 1,$$

which is of smaller order than $k^{-7/8}$ as $k \to \infty$, we can say that

$$\int_0^\infty e^{-kx^8}\sin^6 x\,dx \sim \tfrac{1}{8}k^{-7/8}\,\Gamma(\tfrac{7}{8}) \qquad \text{as } k \to \infty.$$

Widder's results may be generalized in a different direction. The integral may be a double integral, and h and ϕ functions of two variables. The generalization is contained in the following theorem.

(1) *R is a measurable region of the $x - y$ plane, such that if*

$$S_\delta = \{(x, y)\,|\,(x-a)^2 + (y-b)^2 < \delta\},$$

then for some δ_0, $S_{\delta_0} \subset R$,

(2) $\phi(x, y) \in C^2(S_{\delta_0}) \cap L(R)$; $\phi_x = \phi_y = 0$, $\phi_{xy}^2 - \phi_{xx}\phi_{yy} < 0$, $\phi_{xx} < 0$, *at* (a, b); *for each* δ, $0 < \delta < \delta_0$, $\phi(x, y) \leqslant M(\delta) < \phi(a, b)$ *for* $(x, y) \in R - S_\delta$,

(3) $\psi(x, y) \in L(R)$, $\psi(x, y)$ *is continuous at* (a, b) *and* $\psi(a, b) \neq 0$, then

$$\int\int_R \exp\,(k\phi(x, y))\,\psi(x, y)\,dxdy$$

$$\sim 2\pi\,\psi(a, b)\exp\,(k\,\phi(a, b))/k(\phi_{xx}(a, b)\,\phi_{yy}(a, b) - \phi_{xy}^2(a, b))^{\frac{1}{2}}$$

as $k \to \infty$.

Proof. We may assume $\psi(a, b) > 0$. Let

$$x - a = \xi\cos\theta - \eta\sin\theta,\ y = \xi\sin\theta + \eta\cos\theta$$

where

$$\tan 2\theta = 2\phi_{xy}(a, b)/(\phi_{xx}(a, b) - \phi_{yy}(a, b));$$

let

$$\phi(x, y) = \Phi(\xi, \eta),\ \psi(x, y) = \Psi(\xi, \eta).$$

Then since

$$\frac{\partial(x, y)}{\partial(\xi, \eta)} = 1,$$

$$I_k = \int\int_R \exp \{k(\phi(x, y) - \phi(a, b))\} \psi(x, y)dxdy$$
$$= \int\int_R \exp (k(\Phi(\xi, \eta) - \Phi(0, 0)) \Psi(\xi, \eta)\, d\xi d\eta.$$

By 2, 3, and since $\Phi_{\xi\xi}(0, 0) < 0$, $\Phi_{\eta\eta}(0, 0) < 0$, $\Phi_{\xi\eta}(0, 0) = 0$, we can, for every ϵ, $0 < \epsilon < \min. (-\Phi_{\xi\xi}, -\Phi_{\eta\eta}, \Psi(0, 0))$, choose an $\alpha(\epsilon)$, $0 < \alpha < (\frac{1}{2}\delta_0)^{\frac{1}{2}}$, such that,

(i) $A_1 = \Phi_{\xi\xi}(0, 0) - \epsilon < \Phi_{\xi\xi}(\xi, \eta) < B_1 = \Phi_{\xi\xi}(0, 0) + \epsilon < 0,$
$$-\epsilon < \Phi_{\xi\eta}(\xi, \eta) < \epsilon,$$
$$A_2 = \Phi_{\eta\eta}(0, 0) - \epsilon < \Phi_{\eta\eta}(\xi, \eta) < B_2 = \Phi_{\eta\eta}(0, 0) + \epsilon < 0,$$
(ii) $0 < C_1 = \Psi(0, 0) - \epsilon < \Psi(\xi, \eta) < C_2 = \Psi(0, 0) + \epsilon,$

for all (ξ, η), $-\alpha \leqslant \xi \leqslant \alpha$, $-\alpha \leqslant \eta \leqslant \alpha$.

Let $R_\alpha = \{(x, y) | -\alpha \leqslant \xi \leqslant \alpha;\ -\alpha \leqslant \eta \leqslant \alpha\}$. Then

$$I_k = \left\{ \int\int_{R_\alpha} + \int\int_{R-R_\alpha} \right\} \exp \{k(\phi(x, y) - \phi(a, b))\} = I_k' + I_k''.$$

Consider first I_k''. Since, for $(x, y) \in R - R_\alpha$, $(x - a)^2 + (y - b)^2 \geqslant \alpha^2$, we have, by (2),

$$0 \leqslant |I_k''| \leqslant \exp \{k(M(\alpha) - \phi(a, b))\} \int\int_R |\psi(x, y)|\, dxdy$$

so that $I_k'' = O(\omega^k)$ where $0 < \omega < 1$, since $\phi(a, b) > M(\alpha)$, and $\psi(x, y) \in L(R)$.

Now

$$I_k' = \int\int_{R_\alpha} \exp \{k(\phi(x, y) - \phi(a, b))\} \psi(x, y)\, dxdy$$
$$= \int_{-\alpha}^{\alpha}\int_{-\alpha}^{\alpha} \exp \{k(\Phi(\xi, \eta) - \Phi(0, 0))\} \Psi(\xi, \eta)\, d\xi d\eta$$
$$= \left\{ \int_0^{\alpha}\int_0^{\alpha} + \int_{-\alpha}^{0}\int_0^{\alpha} + \int_0^{\alpha}\int_{-\alpha}^{0} + \int_{-\alpha}^{0}\int_{-\alpha}^{0} \right\}$$
$$\exp \{k(\Phi(\xi, \eta) - \Phi(0, 0))\} \Psi(\xi, \eta)\, d\xi d\eta$$
$$= J_1 + J_2 + J_3 + J_4.$$

By (2) and Taylor's theorem,

$$\Phi(\xi, \eta) - \Phi(0, 0) = \tfrac{1}{2}(\Phi_{\xi\xi}(\theta\xi, \theta\eta)\, \xi^2 + 2\Phi_{\xi\eta}(\theta\xi, \theta\eta)\, \xi\eta + \Phi_{\eta\eta}(\theta\xi, \theta\eta)\, \eta^2),$$
$$0 < \theta < 1.$$

Consider J_1. By (i) and (ii),

$$C_1 \int_0^{\alpha}\int_0^{\alpha} \exp (\tfrac{1}{2}k(A_1\xi^2 - 2\epsilon\xi\eta + A_2\eta^2))\, d\xi d\eta < J_1$$
$$< C_2 \int_0^{\alpha}\int_0^{\alpha} \exp (\tfrac{1}{2}k(B_1\xi^2 + 2\epsilon\xi\eta + B_2\eta^2))\, d\xi d\eta.$$

In the left hand side, let $-u^2 = \frac{1}{2}kA_1\xi^2$, $-v^2 = \frac{1}{2}kA_2\eta^2$, and in the right hand side, let $-u^2 = \frac{1}{2}kB_1\xi^2$, $-v^2 = \frac{1}{2}kB_2\eta^2$. Then the inequalities become

$$(2C_1/k(A_1A_2)^{\frac{1}{2}}) \int_0^{k^{1/2}D_1} \int_0^{k^{1/2}D_2} \exp(-(u^2 + 2\epsilon(A_1A_2)^{-\frac{1}{2}} uv + v^2))\, dudv$$

$$< J_1 < (2C_2/k(B_1B_2)^{\frac{1}{2}}) \int_0^{k^{1/2}D_3} \int_0^{k^{1/2}D_4}$$

$$\exp(-(u^2 - 2\epsilon(B_1B_2)^{-\frac{1}{2}} uv + v^2))\, dudv,$$

where $D_i = \alpha(-\frac{1}{2}A_i)^{\frac{1}{2}}$, $(i = 1, 2)$ and $D_i = \alpha(-\frac{1}{2}B_i)^{\frac{1}{2}}$, $(i = 3, 4)$. Hence

$$(2C_1/(A_1A_2)^{\frac{1}{2}}) \int_0^{\infty} \int_0^{\infty} \exp(-(u^2 + 2\epsilon(A_1A_2)^{-\frac{1}{2}} uv + v^2))\, dudv$$

$$\leqslant \overline{\underline{\lim}}\, kJ_1 \leqslant (2C_2/(B_1B_2)^{\frac{1}{2}}) \int_0^{\infty} \int_0^{\infty}$$

$$\exp(-(u^2 + 2\epsilon(B_1B_2)^{-\frac{1}{2}} uv + v^2))\, dudv$$

and thus, since ϵ is arbitrary, and the integrals in question converge uniformly in ϵ for $0 \leqslant \epsilon \leqslant \epsilon_0 < \frac{1}{2}(B_1B_2)^{\frac{1}{2}}$, we have

$$\lim_{k\to\infty} kJ_1 = (2\Psi(0, 0)/(\Phi_{\xi\xi}(0, 0)\Phi_{\eta\eta}(0, 0))^{\frac{1}{2}} \int_0^{\infty} \int_0^{\infty} e^{-(u^2+v^2)}\, dudv$$

$$= \pi\, \psi(a, b)/2(\phi_{xx}(a, b)\, \phi_{yy}(a, b) - \phi_{xy}^2(a, b))^{\frac{1}{2}}.$$

Similarly,

$$\lim_{k\to\infty} kJ_2 = \lim_{k\to\infty} kJ_3 = \lim_{k\to\infty} kJ_4 = \pi\, \psi(a, b)/2(\phi_{xx}(a, b)\, \phi_{yy}(a, b) - \phi_{xy}^2(a, b))^{\frac{1}{2}}.$$

Hence

$$\lim_{k\to\infty} kI_k = \lim_{k\to\infty} kI_k' = \lim_{k\to\infty} k(J_1 + J_2 + J_3 + J_4)$$

$$= 2\pi\, \psi(a, b)/(\phi_{xx}(a, b)\, \phi_{yy}(a, b) - \phi_{xy}^2(a, b))^{\frac{1}{2}}.$$

REFERENCES

1. D. V. Widder, *The Laplace transform* (Princeton, 1946), chap. 7, §§2 and 8.
2. P. S. de Laplace, *Théorie analytique des probabilities* (Paris, 1820), tome 1, partie 2, chap. 1.

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLVII : SERIES III : JUNE, 1953
SECTION THREE

A Remark on Curves of Order n in n-space

P. SCHERK, F.R.S.C.

1. Let R_n denote real projective n-space. An arc A [a curve C] is the continuous image of an interval [of a circle] in R_n. The order of A is the upper limit of the number of points which A may have in common with any [linear] $(n-1)$-space. We exclude the case that A degenerates into a single point. Then the order of A is not less than n. Let A^n [c^n] denote an arc [a curve] of order n in R_n. It necessarily is a Jordan arc [a Jordan curve].

Let $s, s', \ldots$ denote points on the arc A. We call the point s differentiable if it possesses osculating subspaces $L_p{}^n(s)$ of every dimension $[p = -1, 0, 1, \ldots, n]$. Let $L_{-1}{}^n(s)$ be empty space. Suppose we have defined $L_p{}^n(s)$ and postulated its existence. Then we require that the $(p+1)$-spaces through $L_p{}^n(s)$ and a point s' converging to s have a unique limit space. We call it the osculating $(p+1)$-space $L_{p+1}{}^n(s)$. Thus $L_0{}^n(s)$ is the point s itself and $L_n{}^n(s)$ is the whole R_n. We call A differentiable if each of its points is.

It is well known [1] that the end-points of an arc A^n are differentiable. Thus, the interior points of A^n are "one-sidedly differentiable." This implies that A^n is rectifiable and that it is differentiable everywhere with not more than a countable number of points as exceptions.

2. In a recent paper [4], Schoenberg proved with analytic methods that every C^n can be uniformly approximated by analytic C^n's if n is even. He also proved for odd n's that every A^n can be uniformly approximated by arcs of analytic C^n's with the same end-points. In this note, a geometrical proof of the following weaker theorem will be outlined:

Every C^n can be uniformly approximated by differentiable C^n's.

Here n may be any integer greater than one. Our proof is based on [2] and [3]. The author should like to acknowledge the stimulus he received from Dr. Schoenberg and Dr. Motzkin.

3. Given an $(n-1)$-space E and an interior point s of an arc A^n. Let $\bar{A}^n$ be a small closed subarc of A^n which contains s in its interior and which meets E nowhere else. We then call s a point of support

[of intersection] with respect to E if the end-points of $\bar{A}^n$ lie on the same side of E [are separated by E]. The following lemma is quoted from [3]: Suppose s is differentiable and E contains $L_p{}^n(s)$ but not $L_{p+1}{}^n(s)$. If p is odd [even], then s is a point of support [of intersection] with respect to E.

4. Let s and s' be two differentiable points on a curve C^n; $s \neq s'$. They divide C^n into two arcs A^n and A'^n.

The osculating spaces $L_p{}^n(s)$ and $L_{n-p-2}{}^n(s')$ are readily seen to span an $(n-1)$-space which does not meet C^n outside of s and s' and which does not contain $L_{p+1}{}^n(s)$ or $L_{n-p-1}{}^n(s')$ $[-1 \leqslant p \leqslant n-1]$. Altogether there are $n+1$ subspaces of this kind. We verify by induction that they subdivide R_n into 2^n simplexes. From the above, A^n $[A'^n]$ lies entirely in one of these simplexes. The lemma quoted in §3 implies that the simplex S that A'^n lies in is determined by A^n.

Let A''^n be a third arc of order n with the same end-points s and s'. Suppose $A^n \cup A''^n$ is differentiable at s and at s' and shares with C^n the property that the $(n-1)$-space through $L_p{}^n(s)$ and $L_{n-p-2}{}^n(s')$ supports or intersects $A^n \cup A''^n$ at s if p is even respectively odd $[p = 0, 1, 2, \ldots, n-1]$. Then, from [2] and [3], $A^n \cup A''^n$ is a curve of order n. From the above, A''^n will also lie in S. There are differentiable arcs A''^n with the required properties. As a matter of fact, A''^n can be an arc of a rational curve of degree n.

5. Given the curve C^n of order n. We introduce an elliptic metric into R_n and decompose C^n by means of differentiable points into sufficiently small subarcs A'^n. If our decomposition is fine enough, then each of the simplexes S determined by the A'^n will be arbitrarily small. Replacing consecutively each A'^n by a differentiable A''^n, we obtain a curve of order n which is differentiable everywhere and which lies in the union of the simplexes S. This proves our theorem.

Schoenberg's theorem could be proved in this fashion if the following conjecture is true: To any finite set of differentiable points on a C^n there exists an analytic curve of order n through them which has at each of them the same osculating spaces that C^n has.

REFERENCES

1. J. Hjelmslev, *Introduction à la théorie des suites monotones*, D. Kgl. Danske Vidensk. Selsk. Oversigt (1914), no. 1.
2. I. Sauter, Math. Z., vol. 41 (1936), 507–36.
3. P. Scherk, Časopis pro pěst. mat. a fys., vol. 66 (1937), 172–91.
4. I. J. Schoenberg, *On smoothing operations and their generation functions*, Nat. Bur. Standards Rep. no. 1734 (1952).

TRANSACTIONS OF THE ROYAL SOCIETY OF CANADA
VOLUME XLVII : SERIES III : JUNE, 1953
SECTION THREE

The Application of Fourier Transforms in Physical Problems

W. H. WATSON, F.R.S.C.

THE application of Fourier transforms, and integral transforms in general, is of quite modern interest for it is still evolving. Its study exposes in a fundamental way the interactions of mathematics and physics. In the main, books about it suffer from one or other of two faults. Mathematical interest is usually confined to the theory of some special transform such as Mellin or Laplace, and the applications discussed are often trivial, illustrating basic points of theory it is true, but nevertheless falling short as a guide to the detailed analytical technique needed to handle real problems. This afternoon it would be out of place to venture on mathematical elaboration that properly belongs to sustained study of the subject. It is appropriate, however, to recall in mathematical form the basis of integral transforms.

Given a function $K\ (\alpha, x)$ of two variables α and x, and the function $f(x)$, the function $F(\alpha)$ defined by

$$F(\alpha) = \int_0^\infty f(x)\, K(\alpha, x)\, dx$$

is, when the integral is convergent, the integral transform of the function $f(x)$ by the kernel $K\ (\alpha, x)$.

In the Laplace transform $K\ (\alpha, x) = e^{-\alpha x}$
In the Mellin transform $K\ (\alpha, x) = x^{\alpha-1}$
In the Fourier cosine transform $K\ (\alpha, x) = \cos \alpha x$
In the Fourier sine transform $K\ (\alpha, x) = \sin \alpha x$
In the Hankel transform of order ν $K\ (\alpha, x) = (\alpha x)^{\frac{1}{2}}\, J_\nu(\alpha x)$

In each case there is an appropriate inversion formula by which $f(x)$ may be deduced from $F(\alpha)$ [7].

The Fourier transform of $f(x)$ is

$$\bar{f}(\alpha) = \int_{-\infty}^{\infty} f(x)\, e^{i\alpha x}\, dx \tag{1}$$

and the inversion relation is

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \bar{f}(\alpha)\, e^{-i\alpha x}\, d\alpha \tag{2}$$

$f(x)$ must satisfy Dirichlet's conditions and

$$\int_{-\infty}^{\infty} |f(x)|\, dx$$

must converge.

It is forty years since Sommerfeld applied Fourier transformation to treat the propagation of a signal in a dispersive medium. The problem arose from the objection offered to Einstein's Special Relativity theory —an objection supported by citing examples of media in which, since $dV/d\lambda$ is negative, the group velocity exceeds c, and put forward in the belief that the group velocity is necessarily equivalent to the velocity of energy propagation. The particular problem considered by Sommerfeld [8] is that of a signal which originates at $z = 0$, at instant $t = 0$, and continues indefinitely thereafter as a harmonic oscillation of frequency ω_0 (radians/s).

That the Fourier integral of a non-terminating harmonic wave does not converge may be seen from equation (1). The difficulty may be circumvented by deforming the path of integration into the complex domain of the variable ω_0. For example, for

$$f(t) = \begin{cases} \frac{1}{2\pi} e^{-i\omega_0 t} & t > 0 \\ 0 & t > 0, \end{cases}$$

$$\bar{f}(\omega) = \frac{1}{2\pi}\int_0^{\infty} e^{i(\omega-\omega_0)t}\, dt = \frac{1}{2\pi i\,(\omega_0 - \omega)} \tag{3}$$

provided Im $(\omega - \omega_0) > 0$; the integral then converges. Inversely,

$$f(t) = \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{e^{-i\omega t}}{2\pi(\omega_0 - \omega)}\, dt. \tag{4}$$

This integral is evaluated by the theory of residues, the contour being closed by a semicircle at infinity on which, by Jordan's lemma, the contribution to the contour integral is zero. When $t > 0$, the path of integration lies in the negative half-plane at infinity thus enclosing the pole $\omega = \omega_0$. Whereas when $t < 0$ it lies above the real axis and excludes the pole, as shown in Figure 1. This gives the essential discontinuity in the form $f(t)$.

To obtain the Fourier transform of

$$e^{-i\omega_0 t} \qquad (-\infty < t < \infty)$$

we evidently have to combine

$$\bar{f}_+(\omega) = \frac{1}{2\pi}\int_0^\infty e^{i(\omega-\omega_0)t}\,dt \quad\text{and}\quad \bar{f}_-(\omega) = \frac{1}{2\pi}\int_{-\infty}^0 e^{i(\omega-\omega_0)t}\,dt$$

requiring $\mathrm{Im}(\omega - \omega_0) > 0$ for the first and $\mathrm{Im}(\omega - \omega_0) < 0$ for the second. It is in this sense that we write

$$(5)\quad \delta(\omega - \omega_0) = \delta_+(\omega - \omega_0) + \delta_-(\omega - \omega_0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i(\omega-\omega_0)t}\,dt$$

and regard $\delta(\omega - \omega_0)$ as the transform of

$$\frac{1}{2\pi}e^{-i\omega_0 t}.$$

δ has been called the Dirac δ-function. Schwartz [6] calls it the Dirac measure. It is not a function. I shall return to this topic later in the lecture.

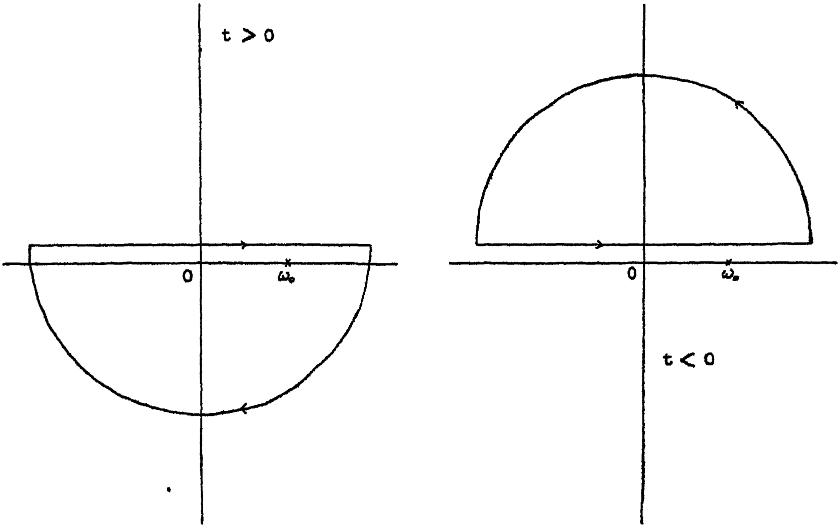


FIGURE 1

The broad use of integral transforms to solve differential equations has developed from Heaviside's operator calculus, devised by him to solve transient problems in physics and electrical engineering. The important core of Heaviside's method is that it corresponds to the possibility of computation with numbers instead of analysis because it starts with what is given, and evolves the answer explicitly by direct operations on the data.

If one wished to characterize the role of Fourier transformation in the theory of physical systems, it may, I think, be fairly stated that Fourier transformation is the natural extension of Fourier series to unbounded spaces permitting the representation of discontinuous distributions in space and time by methods conformable to analysis. Whether discontinuity is a physical necessity does not really matter: it is mathematically simpler than the representation of the actual rapid transition according to any special law that might be physically appropriate, but which, nevertheless, is judged irrelevant to the main idealized purpose of the calculation. This use of discontinuity in representing nature is quite analogous to the use of impulses in dynamics.

It has already been pointed out how the switching on of an oscillation is managed by treating Fourier transformations in the complex domain of the transform variable: a great deal of the subtlety of calculation emerges from this analytical way of representing discontinuous distributions. Since the physicist expects to see in the mathematical development forms of which he recognizes the physical relevance, he will tend to by-pass whatever he thinks he may dismiss as mathematical ritual. This is in some respects very unfortunate, because he is often thereby deprived of the stimulus to appreciate the delicacy of the conceptions being treated and so may not learn to use the very sharp tools of the mathematician.

Let us now make a list of topics in the study of which the application of Fourier transformation has played an important role for theoretical physics.

(a) In the theory of vibrations, Fourier series are the main tool, integral transforms apply usually only to the time dependence.

(b) The diffusion equation and that of heat conduction we might regard as simplified versions of the transport equation and indeed as representing the whole field of continuous linear stochastic theory.

(c) In elasticity stress systems due to discontinuous pressure distributions have been treated.

(d) Hydrodynamical problems involving discontinuous boundary conditions.

(e) Electromagnetic theory and electromechanics have been the field of many successful applications.

(f) Fourier transformation is peculiarly suited for the study of wave equations in general and especially to provide powerful methods for handling singularities and discontinuities.

(g) In modern physics one can point to the use of wave functions in momentum space to treat scattering problems, the use of Mellin and

Laplace transforms in the cascade theories for cosmic rays, and the wide use of Fourier transformation in quantum electrodynamics and wave field theory.

For the mathematician, however, the application of Fourier transformation to physical problems consists of its use to solve (i) partial differential equations and systems of them, (ii) integral equations and systems thereof and in general, (iii) integro-differential equation systems. Of course, Fourier transformation has its role in mathematics itself, and it is an important tool in statistics.

This is an impressive list and naturally not much of it can be exhibited in this lecture. I propose to pick out a few topics to illustrate what has been going on.

The theory of the slowing-down and diffusion of neutrons has received much attention because of its importance for understanding nuclear reactors. We shall consider a very small part of the theory of neutron diffusion under extremely rigorous (indeed unrealistic) physical assumptions designed to simplify the problem [5]. We begin with the transport equation by which the rate of change of the neutron density following the motion is accounted for by capture, scattering and the presence of neutron sources:

$$\frac{d\psi(\mathbf{r}, \beta, t)}{dt} = -P(\beta)\psi(\mathbf{r}, \beta, t) + \int P_s(\beta')f(\beta, \beta')\, d\beta'\psi(\mathbf{r}, \beta', t) + q(\mathbf{r}, \beta, t)$$

with

$$(6) \qquad \frac{d\psi}{dt} = \frac{\partial\psi}{\partial t} + \operatorname{div}(\mathbf{v}\,\psi).$$

The functions ψ and q denote respectively the density of neutrons and of neutron sources both per unit volume at the space position $\mathbf{r}$ at time t, and per unit range of the symbol β that stands for all other parameters required to specify a neutron as, for instance, its velocity $\mathbf{v}$ or direction of motion $\mathbf{S}$. The probabilities per unit time that a neutron is scattered or captured from the range $d\beta$ of β are respectively $P_s(\beta)$ or $P_c(\beta)$. Their sum is $P(\beta)$, the total probability of collision. The relative probability that a neutron originally with parameters β' will be scattered into the range $(\beta', \beta' + d\beta')$ is denoted by $f(\beta', \beta)\, d\beta'$.

The simplifications we shall introduce are:

(a) scattering without change of energy, all neutrons having the same speed v which will be taken as unit;

(b) scattering and sources isotropic;

(c) stationary state.

The mean free path (l) being taken as the unit of length, the unit of time is l/v. Then $P = 1$, $P_c = \alpha$, and $P_s = 1 - \alpha$ and we have in place of (6)

$$\text{(7)} \qquad \mathbf{S}\,.\,\mathrm{grad}\,\psi(\mathbf{r}, \mathbf{S}) + \psi(\mathbf{r}, \mathbf{S}) = \frac{1-\alpha}{4\pi}\,\psi_0(\mathbf{r}) + \frac{1}{4\pi}\,q_0(\mathbf{r})$$

where

$$\psi_0\,(\mathbf{r}) = \int \psi(\mathbf{r}, \mathbf{S}')\,d\mathbf{S}'$$

is the neutron density at $\mathbf{r}$, and $d\mathbf{S}$ is the element of solid angle.

The integro-differential equation (7) to determine ψ is equivalent to an algebraic equation on Fourier transformation, viz.

$$\text{(8)} \qquad -i\mathbf{k}\,.\,\mathbf{S}\,\phi(\mathbf{k}, \mathbf{S}) + \phi(\mathbf{k}\,.\,\mathbf{S}) = \frac{1-\alpha}{4\pi}\,\phi_0(\mathbf{k}) + \frac{1}{4\pi}\,\lambda_0(\mathbf{k})$$

where

$$\phi(\mathbf{k}, \mathbf{S}) = \int \psi(\mathbf{r}, \mathbf{S}) \exp\,(i\mathbf{k}\,.\,\mathbf{r})\,r^2\,dr \sin\theta\,d\theta\,d\phi$$

and $\phi_0(\mathbf{k})$ and $\lambda_0(\mathbf{k})$ are the corresponding integrals for $\psi_0(r)$ and $q_0(r)$. The differential operator S.grad becomes the scalar product.

Now rearrange equation (8) thus

$$\phi(\mathbf{k}, \mathbf{S}) = \frac{1}{4\pi(1 - i\mathbf{k}\,.\,\mathbf{S})}\,\{(1-\alpha)\,\phi_0(\mathbf{k}) + \lambda_0(\mathbf{k})\}.$$

Integrate over all directions S to obtain

$$\text{(9)} \qquad \phi_0(\mathbf{k}) = \{(1-\alpha)\,\phi_0(\mathbf{k}) + \lambda_0(\mathbf{k})\} \int \frac{d\mathbf{S}/4\pi}{1 - i\mathbf{k}\,.\,\mathbf{S}}\,.$$

Since the integral is equal to $k^{-1}\tan^{-1}k$, we may obtain the transform $\phi_0(\mathbf{k})$ of the neutron density directly in terms of the transform $\lambda_0(\mathbf{k})$ of the density of sources:

$$\text{(10)} \qquad \phi_0(\mathbf{k}) = \frac{\tan^{-1}k}{k - (1-\alpha)\tan^{-1}k}\,\lambda_0(\mathbf{k}),$$

which has the form

$$\text{(11)} \qquad \phi_0(\mathbf{k}) = g(k)\,\lambda_0(\mathbf{k}).$$

Because the operations multiplication and convolution correspond on transformation and since $g(k)$ is the Fourier transform of $e^{-r}/4\pi r^2$, the equation (11) reads after inversion

$$\text{(12)} \qquad \psi_0(\mathbf{r}) = \int \frac{e^{-R}}{4\pi R^2}\,q_0(\mathbf{r}')\,d\mathbf{r}',$$

where $R = |\mathbf{r} - \mathbf{r}'|$, showing how the source element at $\mathbf{r}'$ contributes to the neutron density at $\mathbf{r}$. The symbol $d\mathbf{r}'$ denotes the volume element in the space of $\mathbf{r}'$.

As a second problem we shall consider the decay of resonance radiation by spontaneous emission to determine the shape of the resonance line. The physical system consists of one atom initially excited and the radiation field initially devoid of photons. It is represented by a state vector ψ which is a function of the state of the atom and the numbers N_σ of the photons in the states σ. Since we are dealing with resonance radiation we restrict ourselves to the excited state A and the ground state B of the atom and introduce the two relevant basis vectors

$$a(t) \equiv \psi(A, 0_\sigma) \text{ and } b(\nu, t) \equiv \psi(B, 1_\sigma).$$

The fundamental quantum equation

$$\frac{\hbar}{i}\frac{\partial\psi}{\partial t} = H\psi \tag{13}$$

is expressed by the pair

$$\begin{aligned} -i\frac{\partial a}{\partial t} &= \int_0^\infty W(\nu)\, b(\nu, t)\, S(\nu)\, d\nu \\ -i\frac{\partial b}{\partial t} &= W^* a + 2\pi(\nu - \nu_1) b \end{aligned} \tag{14}$$

where $\hbar W(\nu)$ is the interaction operator of the Hamiltonian ν being the photon frequency, and

$$h\nu_1 = E_A - E_B, \qquad S(\nu) = \frac{8\pi^2\nu^2 V}{c^3},$$

and V is the volume of the hohlraum. Equation system (14) is to be solved subject to $a(0) = 1$, $b(\nu, 0) = 0$. On Fourier transformation with respect to t, the transform variable being z (and W^* the complex conjugate of W) we obtain

$$-z\bar{a}(z) = \int_0^\infty W(\nu)\, \bar{b}(\nu, z)\, S(\nu)\, d\nu + a(0) \tag{15a}$$

$$-z\bar{b}(\nu, z) = 2\pi(\nu - \nu_1)\, \bar{b}(\nu, z) + W^*\bar{a}(z). \tag{15b}$$

Solve (15b) for $\bar{b}(\nu, z)$, substitute in (a) and find $a(z)$, which in turn is used to give $\bar{b}(\nu, z)$. On inversion we obtain an integral formula for $b(\nu, t)$, a formal solution from which the shape of the resonance line has been worked out [4].

What I have stated regarding the special connexion of Fourier transformation with discontinuity applies in a peculiarly significant way to the treatment of waves.

For Laplace's equation $\Delta U = 0$ and for the iterated Laplace's equations $\Delta^n U = 0$, which are of elliptic type, there are no solutions other than the analytic functions. However, for a hyperbolic equation such as

$$\frac{\partial^2 U}{\partial x^2} - \frac{\partial^2 U}{\partial y^2} = 0$$

the general solution will be

$$U = A(x + y) + B(x - y)$$

where A and B *may* be continuous functions, not necessarily differentiable, discontinuous functions and even not functions at all in the sense of classical analysis. The introduction of such discontinuous solutions in the theory of partial differential equations is of great importance to physics. It is remarkable that substantially the unorthodoxy initiated by Heaviside is still continuing. The latest adjustment of mathematics to physical intuition is exhibited in the noteworthy Theory of Distributions due to Schwartz. His book on this subject appeared in 1950 and members of this audience who attended the meeting of the Canadian Mathematical Congress at Vancouver in 1949 will recall the excellent series of lectures he delivered on distribution theory. This theory has found a proper mathematical basis for handling discontinuous solutions for partial differential equations and of escaping from the mathematical category of functions to treat Fourier transformation.

For the physicist a constant harmonic wave of infinite duration has an infinitely sharp line spectrum. It is represented by a δ-measure. The inverse relation surely requires the Dirac measure δ to have a Fourier transform. Here physical intuition has kept ahead of mathematics. For

$$\mathscr{F}\delta(\omega - \omega_0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \delta(\omega - \omega_0)\, e^{-i\omega t}\, d\omega = \frac{1}{2\pi} e^{-i\omega_0 t}$$

Now this formal equation we have written down does not fall under the terms of classical analysis, in fact in Lebesgue's theory of integration the integral is zero. Accordingly this formalism can be kept "on the rails" only by physical understanding, whereas the value of mathematics is that logically correct answers are assured by obeying the rules of calculation. Schwartz's work seems likely to improve the situation but there will be a period when notation will be muddled—a matter of serious practical effect in modern theoretical physics.

It is ten years since Heisenberg [3] invented the S-matrix that plays so important a part in quantum electrodynamics and the theory of wave fields. In his papers he represented the asymptotic wave functions, representing for example emergent particles from a reaction, by Fourier transforms. For instance $\delta_+(k_0 - k)$ associated with time dependence $\exp(-ik_0ct)$ represents an outgoing spherical wave as may be seen as follows.

$$\mathcal{F}\delta_+(k_0 - k) = \frac{1}{(2\pi)^3}\int_0^\infty\int_{-1}^{1}\int_0^{2\pi}\frac{e^{-ikr\zeta}}{2\pi i\,(k_0 - k)}\,d\phi\,d\zeta\,k^2dk$$

$$(Im(k_0 - k) > 0)$$

$$= \frac{1}{(2\pi)^3}\,\frac{1}{r}\int_0^\infty\frac{k(e^{ikr} - e^{-ikr})}{(k_0 - k)}\,dk.$$

We assume r large: since k_0 is real Im $k < 0$, and hence the main contribution to the integral is

$$\frac{1}{(2\pi)^3\,r}\int_0^\infty\frac{k\,e^{ikr}}{k_0 - k}\,dk$$

which in turn comes from the vicinity of k_0, the result being

$$\frac{-\,ik_0\,e^{ik_0r}}{4\pi^2r}.$$

In the quantum electrodynamics of Schwinger, Feynman, Dyson *et al.*, one has to deal with the singular wave functions associated with sources and sinks to represent the creation and annihilation of particles. Quite apart from the question of relativistic invariance, the essential mathematical difficulties are concerned with the fact that the elementary solutions of the various wave equations (due to point singularities) are properly speaking not functions at all. Nevertheless one writes down expressions for their Fourier transforms.

Fourier transformation enters as the obvious means for dealing with the convolution integrals that appear in the calculation of the wave functions representing the quantum stochastic process [1]. For example, a particle starts from A travels to B where it is scattered, what is the probability amplitude to reach C from A after such a single scattering process? It has the form

$$\int g(\mathbf{x} - \mathbf{x}')\,f(\mathbf{x}')\,d\mathbf{x}'$$

and its Fourier transform is $G(k)F(k)$ which is a straightforward product. The transforms are all singular and it is in the handling of them by conventional means that some of the difficult divergencies in the theory arise. In a recent paper, Güttinger [2] has shown how Schwartz's Distribution Analysis automatically leaves a place for mathematical operations equivalent in effect to renormalization, regularization, or cut-off procedures: it does this quite naturally and brings to light that the theory as at present conceived is physically incomplete. There is a good expectation that distribution analysis may remove some of the mathematical difficulties of quantum field theory especially as applied to mesons.

Thus we view today an interesting prospect. New mathematical tools have become available to the theoretical physicist, but perhaps even more promising is the realization that mathematicians have been too ready in the past to impose on physics the unnecessarily restrictive categories of their own inventions. It is significant that the application of Fourier transforms has been the source of these developments that illuminate the role of mathematics in relation to physics.

REFERENCES

1. R. P. Feynman, Rev. Mod. Phys., vol. 20 (1948), 367–87: Phys. Rev., 76, 749 759, 1949.
2. W. Güttinger, Phys. Rev., vol. 89 (1953), 1004.
3. W. Heisenberg, Z. Phys., vol. 120 (1943), 513-38, 673.
4. E. J. Hellund, Phys. Rev., vol. 89 (1953), 919.
5. G. Placzek and G. M. Volkoff, *Notes on diffusion of neutrons without change in energy,* National Research Council of Canada, MT4, April 1943.
6. L. Schwartz, *Théorie des distributions*, tomes I et II (Paris, Hermann et Cie, 1950).
7. I. N. Sneddon, *Fourier transforms* (New York, McGraw-Hill, 1951).
8. J. Stratton. *Electromagnetic theory* (New York, McGraw-Hill, 1941), chap. v.